

On Asymptotic Incoherence and its Implications for Compressed Sensing of Inverse Problems

A. D. Jones, B. Adcock and A. C. Hansen

Abstract

Recently, it has been shown that incoherence is an unrealistic assumption for compressed sensing when applied to many inverse problems. Instead, the key property that permits efficient recovery in such problems is so-called local incoherence. Similarly, the standard notion of sparsity is also inadequate for many real world problems. In particular, in many applications, the optimal sampling strategy depends on asymptotic incoherence and the signal sparsity structure. The purpose of this paper is to study asymptotic incoherence and its implications towards the design of optimal sampling strategies and efficient sparsity bases. It is determined how fast asymptotic incoherence can decay in general for isometries. Furthermore it is shown that Fourier sampling and wavelet sparsity, whilst globally coherent, yield optimal asymptotic incoherence as a power law up to a constant factor. Sharp bounds on the asymptotic incoherence for Fourier sampling with polynomial bases are also provided. A numerical experiment is also presented to demonstrate the role of asymptotic incoherence in finding good subsampling strategies.

1 Introduction

Compressed sensing, introduced by Candès, Romberg & Tao [10] and Donoho [15], has been one of the major achievements in applied mathematics in the last decade [8, 14, 16, 17, 18]. By exploiting additional structure such as sparsity and incoherence, one can solve inverse problems by uniform random subsampling and convex optimisation methods, and thereby recover signals and images from far fewer measurements than conventional wisdom suggests.

However, in many applications – including Magnetic Resonance Imaging (MRI) [19, 27], X-ray Computed Tomography [11, 29], Electron Microscopy [25, 26], etc – incoherence is always lacking if one tries to keep the model in its original continuous form, as can be seen from Figure 1. The reason for this can be traced to the observation that many classical inverse problems are based on continuous integral transforms such as the Fourier transform:

$$g = \mathcal{F}f, \quad f \in L^2(\mathbb{R}^d), \quad \mathcal{F}f(\omega) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \omega \cdot x} dx. \quad (1)$$

In this case the resulting recovery problem is that of reconstructing an unknown function f from pointwise samples of g .

In compressed sensing, such a transform is combined with an appropriate sparsifying transformation associated to a basis or frame, giving rise to an infinite *measurement* matrix. The *coherence* of an infinite matrix¹ $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ or a finite matrix $W \in \mathbb{C}^{N \times N}$ is defined as

$$\mu(U) = \max_{i,j \in \mathbb{N}} |U_{ij}|^2, \quad \mu(W) = \max_{i,j=1,\dots,N} |W_{ij}|^2. \quad (2)$$

In the finite-dimensional case, the matrix W is typically a change of basis matrix from some *Sampling Basis* $(\rho(m))_{m=1}^N$ to some *Reconstruction Basis* $(\tau(m))_{m=1}^N$ of \mathbb{C}^N :

$$W_{m,n} = \langle \tau(n), \rho(m) \rangle. \quad (3)$$

¹The notation $\mathcal{B}(\ell^2(\mathbb{N}))$ refers to the space of bounded linear maps from $\ell^2(\mathbb{N})$ to itself.

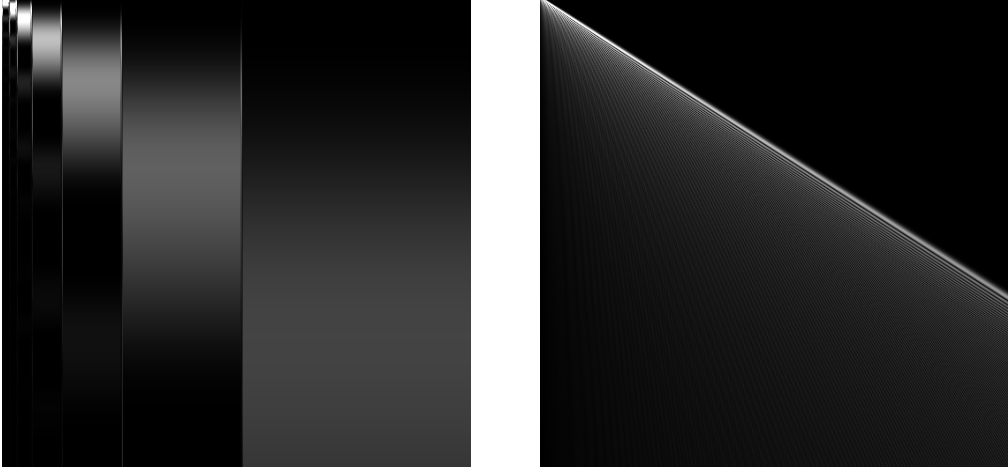


Figure 1: Plots of the absolute values of the entries of the compression matrix U corresponding to Fourier sampling with Daubechies6 boundary wavelets (left) and Legendre polynomials (right). Light regions correspond to large values and dark regions to small values. $U_{1,1}$ corresponds to the top-left entry. Notice that if these matrices were incoherent, then these figures would have been monotone. This means an alternative to coherence is needed to study infinite dimensional compression matrices.

Similarly for the infinite-dimensional case, the matrix U is instead formed from a sampling basis $(\rho(m))_{m \in \mathbb{N}}$ and reconstruction basis $(\tau(m))_{m \in \mathbb{N}}$ of $L^2(\mathbb{R}^d)$:

$$U_{m,n} = \langle \tau(n), \rho(m) \rangle. \quad (4)$$

The term *incoherence* refers to $\mu(U), \mu(W)$ being small. In the finite case, if we assume W is an isometry with respect to the Euclidean norm, then the statement that W is incoherent can be interpreted as W being evenly flat or spread out. In the limiting case where $\mu(W) = N^{-1}$, every entry must have the same absolute size and the matrix is perfectly flat. In the infinite case, such a notion of uniform flatness is impossible for an isometry $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, since its columns are also infinite and normalised.

In the search for an alternative, examples often guide the way. Wavelets, or their various generalizations, are frequently used as the sparsifying transformation, and for smooth functions, one often considers orthogonal polynomials. As Figure 1 reveals, although such a measurement matrix is clearly not incoherent, it is *asymptotically incoherent*: that is to say, the high coherence terms (large matrix entries) are concentrated around a submatrix of U .

Therefore one should not only consider the notion of the coherence of a sensing matrix U , but also a notion of *local coherence*, which we *define* as the coherence over a submatrix of U . One way to generate local coherences is by using the following projection operators $\pi_N, R_N : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$

$$\pi_N(x)_i := \begin{cases} 0 & i \neq N \\ x_i & i = N \end{cases}, \quad R_N(x)_i := \begin{cases} 0 & i < N \\ x_i & i \geq N \end{cases}, \quad (5)$$

and look at the decay of the corresponding *line and block coherences*

$$\mu(\pi_N U), \mu(U \pi_N), \mu(R_N U), \mu(U R_N). \quad (6)$$

The term ‘Line coherence’ refers to $\mu(\pi_N U)$ ($\mu(U \pi_N)$) being equal to the squared absolute maximum of the N th row (column) of U . Likewise $\mu(R_N U)$ ($\mu(U R_N)$) is equal to the squared absolute maximum over U without the first $N - 1$ rows (columns). ‘Asymptotic incoherence’ refers to the decay of the line/block coherences as $N \rightarrow \infty$.

Notions of local coherence have been studied before [5, 4, 23]. For example, in [23] the “local coherence” between two bases $\Phi = (\varphi_j)_{j=1, \dots, N}$, $\Psi = (\psi_k)_{k=1, \dots, N}$ of \mathbb{C}^N was defined as

$$\mu_j^{loc}(\Phi, \Psi) = \sup_{1 \leq k \leq N} |\langle \varphi_j, \psi_k \rangle|, \quad j = 1, \dots, N. \quad (7)$$

This is analogous to a discrete version of the line coherence $\mu(\pi_N U)$.

It is also worth mentioning that lack of coherence is not the only problem that people are faced with when trying to model continuous problems of Fourier type. For example in NMR based problems such as MRI samples are not taken pointwise but in paths or lines [7], restricting the ability to subsample freely.

1.1 Order Notations and Conventions

Throughout this paper we will be using the following notations and conventions to succinctly describe various types of decay; for $f, g : \mathbb{N} \rightarrow \mathbb{R}_{>0}$, $0 \notin \mathbb{N}$:

$$\begin{aligned} f(N) = \mathcal{O}(g(N)) &\Leftrightarrow \exists C > 0 \text{ s.t. } f(N) \leq C \cdot g(N) \quad \forall N \in \mathbb{N}, \\ f(N) = o(g(N)) &\Leftrightarrow \forall C > 0 \quad f(N) \leq C \cdot g(N) \quad \forall N \in \mathbb{N}, \\ f(N) = \Theta(g(N)) &\Leftrightarrow \exists C_1, C_2 > 0 \text{ s.t. } C_1 \cdot g(N) \leq f(N) \leq C_2 \cdot g(N) \quad \forall N \in \mathbb{N}. \end{aligned} \quad (8)$$

Moreover for $f, g : S \rightarrow \mathbb{R}_{>0}$ where S is a set we write

$$f \lesssim g \Leftrightarrow \exists C > 0 \text{ s.t. } f(s) \leq C \cdot g(s) \quad \forall s \in S. \quad (9)$$

1.2 Compressed Sensing and the Coherence Barrier

Let us now provide some background regarding compressed sensing and incoherence.

Working in the finite dimensional case, standard compressed sensing theory [9, 3] says that if $x \in \mathbb{C}^N$ is s -sparse, i.e. x has at most s nonzero components, then, with probability exceeding $1 - \epsilon$, x is the unique minimiser to the problem

$$\min_{\eta \in \mathbb{C}^N} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega W \eta = P_\Omega W x, \quad (10)$$

where P_Ω is the projection onto $\text{span}\{e_j : j \in \Omega\}$, $\{e_j\}$ is the canonical basis, Ω is chosen uniformly at random with $|\Omega| = m$ and

$$m \gtrsim \mu(W) \cdot N \cdot s \cdot \log(\epsilon^{-1}) \cdot \log(N), \quad (11)$$

The estimate 11 demonstrates how the three pillars of compressed sensing – sparsity, incoherence and uniform random subsampling – combine to allow for recovery with substantial subsampling. Coherence-based sampling has proved effective in a number of different CS scenarios, e.g. polynomial interpolation [20]. However, now suppose that $\mu(W)$ is large; for example, $\mu(W) \cdot N = \mathcal{O}(N)$ as $N \rightarrow \infty$. In this case, 11 suggests that no dramatic subsampling is possible: that is, we must take roughly N samples to recover x , even though x is often extremely sparse. We refer to this phenomenon as the *coherence barrier*.

1.3 The flip test: How to get the model right

The discussion above and Figure 1 suggests that incoherence may not be the appropriate tool to describe compressed sensing when used in infinite dimensional problems involving basis transforms, as is the case with MRI using Fourier samples. Even though there has already been some progress on developing sampling methods dependent on local coherence and conventional sparsity [23], it turns out sparsity must be revised as well. The flip test is a numerical tool designed to verify what kind of sparsity structure we actually recover in practical applications such as MRI, CT, Electron microscopy etc.

The initial test was first introduced in [4], and we will present variants of it here. The first flip test was designed to answer the following question: Is the success of subsampling techniques used in

compressed sensing independent of the location of the coefficients to be recovered? In other words is sparsity the right model for compressed sensing?

The flip test can be described as follows Let $x \in \mathbb{C}^N$ be a vector, and $U \in \mathbb{C}^{N \times N}$ a measurement matrix. We then sample according to some pattern $\Omega \subseteq \{1, \dots, N\}$ with $|\Omega| = m$ and solve (10) for x , i.e. $\min \|z\|_1$ s.t. $P_\Omega U z = P_\Omega U x$ to obtain a reconstruction $z = \alpha$. Now we flip x to obtain a vector x' with reverse entries, $x'_i = x_{N-i}$, $i = 1, \dots, N$ and solve (10) for x' using the same U and Ω , i.e. $\min \|z\|_1$ s.t. $P_\Omega U z = P_\Omega U x'$. Assuming z to be a solution, then by flipping z we obtain a second reconstruction α' of the original vector x , where $\alpha'_i = z_{N-i}$. If the success of the sampling technique is independent of the structure of the coefficients, then we should have that α and α' are close. As Figure 2 suggests, this is not the case.

Another version of the flip test is to permute the coefficients in the wavelet levels that corresponds to the different scales. In other words, instead of flipping the coefficients completely, the coefficients are permuted only within a certain level, but never across the levels. The result of this is visualised in Figure 2. Note that the results of the flip test presented here carry over to different subsampling schemes as well (see [2]).

Conclusion of the flip tests:

- (i) The optimal sampling strategy depends on the signal structure.
- (ii) There is no uniform recovery, hence the matrix involved does not satisfy the Restricted Isometry Property (RIP).
- (iii) Theories based on sparsity will not explain the success of the recovery. Instead a notion of sparsity structure where there is a sparsity s_k of important coefficients in the k th wavelet level is needed.
- (iv) Because of this dependence on sparsity structure, both local coherences in the sampling basis $\mu(\pi_N U)$ and the sparsity basis $\mu(U \pi_N)$ must be considered.

Remark 1. *The flip test can be extended to show [2] that standard successful sampling schemes used in MRI, Electron Tomography, Neutron/ ^3He Scattering, Fluorescence Microscopy etc. do not recover all weighted sparse [2] vectors when using wavelets (or other X-lets). In particular, weighted sparsity suffers from similar issues as conventional sparsity, namely that the class is too big and allows for vectors that cannot be recovered with standard sampling schemes. Hence weighted sparsity is insufficient for modelling compressed sensing with wavelets. However, theory and tests demonstrate that sparsity in levels (as in (iii)) may be a much more realistic model for compressed sensing. Motivated by this we use that as the structured sparsity model in this paper.*

1.4 Overcoming the Coherence Barrier

When faced with the coherence barrier, the standard compressed sensing approach of subsampling uniformly at random does not work. This begs the question: do we have an alternative? Empirically, it is known that the answer to this question is yes: one can break the coherence barrier by sampling at different rates over different frequency ranges. This was recently confirmed by mathematical analysis in [5, 4]. The key to their work was to replace the three principles of compressed sensing with the three concepts of *sparsity in levels*, *multi-level sampling* and *local coherence* – and prove recovery estimates akin to 11 under these more general settings.

Definition 2 (Sparsity in levels). *Let x be an element of either \mathbb{C}^N or $l^2(\mathbb{N})$. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \dots < M_r$ and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \dots, r$, where $M_0 = 0$. We say that x is (\mathbf{s}, \mathbf{M}) -sparse if, for each $k = 1, \dots, r$,*

$$\Delta_k := \text{supp}(x) \cap \{M_{k-1} + 1, \dots, M_k\},$$

satisfies $|\Delta_k| \leq s_k$. We denote the set of (\mathbf{s}, \mathbf{M}) -sparse vectors by $\Sigma_{\mathbf{s}, \mathbf{M}}$.

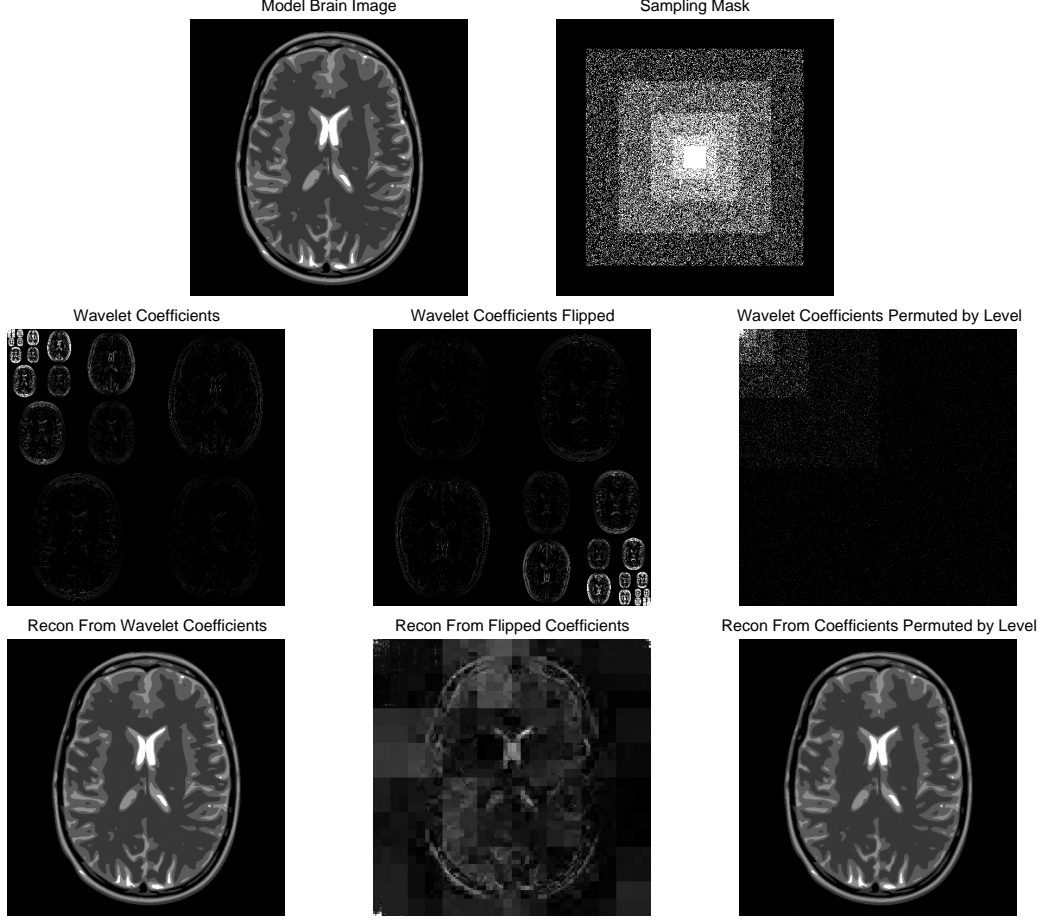


Figure 2: A standard flip test demonstrating the importance of the location of wavelet coefficients on the success of CS subsampling techniques. Here we decompose the Brain image into a two-dimensional separable Haar basis and solve for these coefficients through subsampling Fourier data of the corresponding image (Resolution= 512×512). We then compare the results to the same wavelet data but the order flipped versus randomly permuting the coefficients inside each wavelet level. After using CS to solve for the flipped/permuted coefficients, the data is then flipped/permuted back and then imaged.

Definition 3 (Multi-level sampling scheme). *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, and suppose that*

$$\Omega_k \subseteq \{N_{k-1} + 1, \dots, N_k\}, \quad |\Omega_k| = m_k, \quad k = 1, \dots, r,$$

are chosen uniformly at random, where $N_0 = 0$. We refer to the set

$$\Omega = \Omega_{\mathbf{N}, \mathbf{m}} := \Omega_1 \cup \dots \cup \Omega_r$$

as an (\mathbf{N}, \mathbf{m}) -multilevel sampling scheme. Observe that such sampling schemes are different to that of taking independent identically distributed (IID) samples from a common law.

Definition 4. *Let U be an isometry of either \mathbb{C}^N or $l^2(\mathbb{N})$. If $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$ and $1 \leq M_1 < \dots < M_r$. In [4] the authors introduced their own definition of local coherence as follows; the $(k, l)^{\text{mathrm{th}}}$ local coherence of U with respect to \mathbf{N} and \mathbf{M} is*

$$\mu_{\mathbf{N}, \mathbf{M}}(k, l) = \sqrt{\mu(P_{N_{k-1}}^{N_k} U P_{M_{l-1}}^{M_l}) \cdot \mu(P_{N_{k-1}}^{N_k} U)}, \quad k, l = 1, \dots, r, \quad (12)$$

where $N_0 = M_0 = 1$ and P_b^a denotes the projection matrix corresponding to the indices $\{a + 1, \dots, b\}$.

In [4] a new theory of compressed sensing for changes of bases between infinite-dimensional Hilbert-spaces was introduced based on these assumptions. In this case we solve the following problem: if $x \in \ell^2(\mathbb{N})$ is (s, \mathbf{M}) -sparse and U is an isometry of $\ell^2(\mathbb{N})$ then we hunt for the (hopefully unique) η that solves

$$\min_{\eta \in \ell^2(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U \eta = P_\Omega U x, \quad (13)$$

In this case, instead of a standard compressed sensing estimate 11 determining the total number of measurements, one has the following estimate regarding the local number of measurements m_k in the $k^{\text{mathrm{th}}$ level:

$$m_k \gtrsim (N_k - N_{k-1}) \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^r \mu_{\mathbf{N}, \mathbf{M}}(k, l) \cdot s_l \right) \cdot \log(N), \quad k = 1, \dots, r. \quad (14)$$

When the above condition is satisfied, [4] found that with probability exceeding $1 - \epsilon$, x can be shown to be the unique minimiser to (13). In particular, the number of samples m_k needed to be taken in each region $\{N_{k-1} + 1, \dots, N_k\}$ can be inferred through the local sparsities and coherences using the asymptotic relation (14).

1.5 Using Asymptotic Incoherence to Infer How to Sample

This estimate begs the following question: how do the local sparsity and incoherences behave in practice? As described in [5, 4], natural images possess not just sparsity, but so-called asymptotic sparsity. That is, the ratios $s_k/(N_k - N_{k-1}) \rightarrow 0$ as $k \rightarrow \infty$ in any appropriate basis (e.g. wavelets and their generalizations).

We also observe that the coherence term $\mu_{\mathbf{N}, \mathbf{M}}(k, l)$ can be estimated as follows:

$$\mu_{\mathbf{N}, \mathbf{M}}(k, l) \leq \sqrt{\min(\mu(R_{N_{k-1}+1}U), \mu(UR_{M_{l-1}+1})) \cdot \mu(R_{N_{k-1}+1}U)}. \quad (15)$$

Therefore, combining this with (14), it is clear from 15 that in order to determine the appropriate sampling density one needs good estimates for the block coherences $\mu(R_N U)$ and $\mu(UR_N)$ for $N \in \mathbb{N}$. This is the key contribution of this paper.

1.6 Related Results

Before stating our main results we will discuss the related work in [23, 22, 30]. A key point regarding these results is that they are based on the sparsity model and the Restricted Isometry Property (RIP). It is clear from the flip test that sparsity is not the right model for the Fourier to wavelet case which is central to MRI and numerous other applications. In particular, sampling strategies that seek to recover sparse vectors will have to take an unrealistically large number of samples. This is also reflected in the theoretical guarantees of [23] which exhibit several additional logarithmic factors over traditional compressed sensing estimates. This is in contrast to the framework based on sparsity in levels [4] (which is the basis of this paper) that is shown both empirically [31] and theoretically [2] to even outperform incoherent sampling such as random sub-Gaussian, permuted Fourier, expanders etc.

In [23], they work with discretised Fourier and Haar wavelet elements φ_k and $h_{n,l}^e$ defined on $\mathbb{N}^N = \mathbb{N}^{2^p}$ as follows:

$$\begin{aligned} \varphi_k(t) &= \frac{1}{\sqrt{N}} e^{2\pi i t k / N}, \quad -N/2 + 1 \leq k \leq N/2 \in \mathbb{Z}, \\ h^0(t) &= 2^{-p/2}, \quad h^1(t) = \begin{cases} 2^{-p/2}, & 1 \leq t \leq 2^{p-1}, \\ -2^{-p/2}, & 2^{p-1} < t \leq 2^p, \end{cases} \\ h_{n,l}^e(t) &= 2^{n/2} h^e(2^n t - 2^p l), \quad 0 < n < p, \quad 0 \leq l < 2^n. \end{aligned} \quad (16)$$

Using this notation the following upper bound on the discretised 1D Fourier/Haar wavelet case were derived (Lemma 6.1):

$$|\langle \varphi_k, h_{n,l} \rangle| \leq \min \left(\frac{6 \cdot 2^{n/2}}{|k|}, 3\pi 2^{-n/2} \right), \quad k = -N/2 + 1, \dots, N/2 \in \mathbb{Z}, \quad n = 1, \dots, p-1 \in \mathbb{N}, \quad (17)$$

Result (17) was used to derive the following bound in Corollary 6.4:

$$|\langle \varphi_k, h_{n,l} \rangle| \leq \frac{3\sqrt{2\pi}}{\sqrt{k}}, \quad k = -N/2 + 1, \dots, N/2 \in \mathbb{Z}. \quad (18)$$

With the terminology of our paper, the result (18) shows that the corresponding discrete change of basis matrix $U_N \in \mathbb{C}^N \times \mathbb{C}^N$ satisfies (ordering the φ_k by frequency c.f Def. 13)

$$\mu(\pi_N U_N) = \mathcal{O}(N^{-1}). \quad (19)$$

The closest results to (19) in this paper are for a continuous change of basis matrix between Fourier elements and wavelets in $L^2[-1, 1]$ shown in (23)

$$\mu(\pi_N U) = \Theta(N^{-1}), \quad \mu(U \pi_N) = \Theta(N^{-1}). \quad (20)$$

This result covers *all Daubechies wavelet bases* and not just the Haar case. Moreover the result provides asymptotic lower bounds.

The methods of proof are also worth comparing. In [23] explicit forms of the discrete Haar basis in Section 2.2 are used to derive (18). For higher order Daubechies wavelets such an explicit description is not available and this approach cannot be extended. In this paper we instead rely on the fact that the entries of U can be viewed as Fourier transforms of the wavelet basis e.g. (31). Since Daubechies wavelets were originally constructed through their Fourier transform, this analytical approach is very natural.

This work is also mentioned in [30] along with further results for pointwise sampling of polynomials with a weighted sparsity model. As discussed in [2], weighted sparsity suffers from the same issues as sparsity and becomes an unrealistic model for Fourier samples and wavelet recovery. Moreover, the results in this paper on polynomials are with Fourier sampling which has very little connections with direct point samples considered in [30].

2 Main Results

2.1 Coherence Bounds

Our main results provide estimates for the precise convergence rates of $\mu(P_N U)$ and $\mu(U P_N)$ in the case of one-dimensional Fourier-wavelet and Fourier-polynomial bases spanning $L^2[-1, 1]$. For $\epsilon \in (0, 1/2]$ fixed, the Fourier basis $B_f(\epsilon)$ is defined as

$$\chi_k(x) = \sqrt{\epsilon} \exp(2\pi i \epsilon k x) \cdot \mathbb{1}_{[(-2\epsilon)^{-1}, (2\epsilon)^{-1}]}(x), \quad x \in \mathbb{R}, \quad k \in \mathbb{Z}. \quad (21)$$

Notice that because $\epsilon \in (0, 1/2]$, $B_f(\epsilon)$ is a basis of $L^2[(2\epsilon)^{-1}, (2\epsilon)^{-1}]$. The standard wavelet basis B_w for a given Daubechies scaling function ϕ and wavelet ψ consists of functions of the form

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k). \quad (22)$$

A more precise description of these bases is provided in Section 6. Suppose $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ is the change of basis matrix formed by the pair of bases $(B_f(\epsilon), B_w)$ (see Definition 28 for a formal definition of U). Since $U : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ the bases must be indexed by \mathbb{N} which means we must enumerate the bases in some way using orderings (see Section 4). For $B_f(\epsilon)$ we enumerate with increasing frequency (a *frequency ordering*) and for B_w we order with j increasing (a *leveled ordering*).

Theorem 5 (Fourier-Wavelet Case). *Under the above conditions (with $\epsilon \in I_{J,p}$, see Remark 16), we have*

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1}). \quad (23)$$

From Lemma 11 we also deduce that $\mu(R_N U), \mu(U R_N) = \Theta(N^{-1})$.

Next let $B_p = (\tilde{p}_n)_{n \in \mathbb{N}}$ denote the basis of L^2 -normalised Legendre polynomials on $[-1, 1]$. Suppose $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ now denotes the change of basis matrix formed by the pair of bases $(B_f(\epsilon), B_p)$ with a frequency ordering of $B_f(\epsilon)$ (B_p is already ordered by polynomial degree).

Theorem 6 (Fourier-Polynomial Case). *In this case (given $\epsilon \in (0, 0.45]$), we have*

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-2/3}), \quad (24)$$

and we deduce $\mu(R_N U), \mu(U R_N) = \Theta(N^{-2/3})$.

Theorems (5) and (6) are covered by Corollaries 22 and 26 respectively.

These results suggest that subsampling using compressed sensing is in general more effective for the Fourier-wavelet case than the Fourier-polynomial case, assuming similar sparsity structures.

2.2 Optimality Results

We also show that (23) and (24) cannot be improved by changing the orderings of the two bases.

Theorem 7 (Optimality). *For the Fourier-wavelet case $(B_f(\epsilon), B_w)$, none of the following decay rates can be achieved*

$$\mu(R_N U) = o(N^{-1}), \quad \mu(U R_N) = o(N^{-1}). \quad (25)$$

For the Fourier-polynomial case $(B_f(\epsilon), B_p)$, the following decay rates are impossible no matter what orderings of the bases are used:

$$\mu(R_N U) = o(N^{-2/3}), \quad \mu(U R_N) = o(N^{-2/3}). \quad (26)$$

Theorem 33 covers this result.

Finally we look at the general case we only impose that $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ is an isometry and ask how fast $\mu(R_N U)$ can possibly decay. Theorem 34 states that

$$\sum_{N \in \mathbb{N}} \mu(R_N U) < \infty, \quad (27)$$

must hold in general. Therefore $\mu(R_N U) = o(N^{-\alpha})$ is impossible for $\alpha > 1$ showing that the Fourier-wavelet decay (23) attains the fastest theoretically possible decay rate as a power law. Furthermore Theorem 36 shows that, as a statement for all isometries $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, (27) cannot be improved upon.

3 Outline for the Remainder of the Paper

This outline is for those wishing to quickly extract the results and proofs in this paper. Section 4 is mandatory reading for all the other sections. Section 5 is also important for understanding the bases and orderings needed for the one-dimensional coherence bounds. Sections 6 and 7 covering the Fourier-wavelet Fourier-polynomial cases are independent of each other. Section 8 demonstrates how these two cases work in practice with some numerical data. Sections 9 and 10 cover optimality results and theoretical limits and are independent of Sections 5-8, barring the applications of Theorem 33 to (23) and (24) which is used as motivation. Finally in Section 11 we discuss an alternative to the notion of optimality presented in Section 9.

4 Coherences and Orderings

We work in an infinite dimensional separable Hilbert space \mathcal{H} with two closed infinite dimensional subspaces V_1, V_2 spanned by orthonormal bases B_1, B_2 respectively,

$$V_1 = \overline{\text{Span}\{f \in B_1\}}, \quad V_2 = \overline{\text{Span}\{f \in B_2\}}.$$

We call (B_1, B_2) a ‘basis pair’.

Definition 8 (Orderings). *Let S be a set. Say that a function $\rho : \mathbb{N} \rightarrow S$ is an ‘ordering’ of S if it is bijective.*

Definition 9 (Change of Basis Matrix). *For a basis pair (B_1, B_2) , with corresponding orderings $\rho : \mathbb{N} \rightarrow B_1$ and $\tau : \mathbb{N} \rightarrow B_2$, form a matrix $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ by the equation*

$$U_{m,n} := \langle \tau(n), \rho(m) \rangle. \quad (28)$$

Whenever a matrix U is formed in this way we write $\mathcal{U} := [(B_1, \rho), (B_2, \tau)]$.

At this point it is wise to look at how the orderings of the two bases effect the various notions of local coherence.

Lemma 10. *Let $U = [(B_1, \rho), (B_2, \tau)]$. For the coherence terms with the projection on the left hand side, i.e. $\mu(R_N U), \mu(\pi_N U)$, there is no dependence on the choice of ordering of the second basis τ . Likewise the coherences $\mu(U R_N), \mu(U \pi_N)$ do not depend upon the ordering of the first basis ρ .*

This result follows immediately from the definitions of the line/block coherences (5). Even though $\mu(R_N U)$ does not depend on τ , dependence on ρ is so strong that arbitrarily slow decay of $\mu(R_N U)$ is possible for any isometry $U \in \ell^2(\mathbb{N})$ by varying ρ .

Next we observe that bounds on the line coherences translates into bounds for the corresponding block coherences.

Lemma 11. *Let $U = [(B_1, \rho), (B_2, \tau)]$. Suppose $\mu(\pi_N U) = \Theta(f(N))$ for some decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$. Then $\mu(R_N U) = \Theta(f(N))$. Likewise for $\mu(U \pi_N)$ and $\mu(U R_N)$.*

Proof. The lower bound is immediate since $\mu(R_N U) \geq \mu(\pi_N U)$ by definition. The upper bound follows by observing that

$$\mu(R_N U) = \max_{M \geq N} \mu(\pi_M U) \leq C_2 \max_{M \geq N} f_2(M) = C_2 f(N). \quad (29)$$

□

Throughout this paper we would like to define an ordering according to a particular property of a basis but this property may not be enough to specify a unique ordering. To deal with this issue we introduce the notion of consistency:

Definition 12 (Consistent ordering). *Let $F : S \rightarrow \mathbb{R}$ where S is a set. We say that an ordering $\rho : \mathbb{N} \rightarrow S$ is ‘consistent with respect to F ’ if*

$$F(f) < F(g) \quad \Rightarrow \quad \rho^{-1}(f) < \rho^{-1}(g), \quad \forall f, g \in S.$$

5 Bases & Ordering

5.1 Fourier Basis

We recall the definition of the Fourier basis $B_f(\epsilon)$ from (21).

Definition 13 (Frequency ordering). *We define $F_f : B_f \rightarrow \mathbb{N} \cup \{0\}$ by $F_f(\chi_k) = |k|$ and say that an ordering $\rho : \mathbb{N} \rightarrow B_f$ is a ‘frequency ordering’ if it is consistent with F_f .*

For convenience in what follows we shall identify $B_f(\epsilon)$ with \mathbb{Z} by the function $\lambda : B_f \rightarrow \mathbb{Z}$, $\lambda(\chi_k) := k$ which means that for any ordering ρ of $B_f(\epsilon)$ we have

$$\rho(m)(x) = \sqrt{\epsilon} \exp(2\pi i \epsilon \cdot \lambda \circ \rho(m)x) \cdot \mathbb{1}_{[(-2\epsilon)^{-1}, (2\epsilon)^{-1}]}(x), \quad \forall m \in \mathbb{N}.$$

Definition 13 says that an ordering ρ of $B_f(\epsilon)$ is a frequency ordering if and only if the function $|\lambda \circ \rho|$ is nondecreasing. Therefore ρ is a frequency ordering if and only if we have $\{\lambda \circ \rho(2n), \lambda \circ \rho(2n+1)\} = \{+n, -n\}$ for $n \in \mathbb{N}$ and $\lambda \circ \rho(1) = 0$ and consequently $|\lambda \circ \rho(m)| = \lceil (m-1)/2 \rceil$.

5.2 Legendre Bases

If $(R_N)_{n \in \mathbb{N}}$ denotes the standard Legendre polynomials on $[-1, 1]$ (so $R_N(1) = 1$) then the L^2 -normalised Legendre polynomials are defined by $\tilde{p}_n = \sqrt{n-1/2} \cdot R_N$ and we write $B_p := (\tilde{p}_n)_{n=1}^\infty$. The basis B_p is already ordered; call this the *natural ordering*.

5.3 Standard Wavelets

Take a Daubechies wavelet ψ and corresponding scaling function ϕ in $L^2(\mathbb{R})$ with

$$\text{Supp}(\phi) = \text{Supp}(\psi) = [-p+1, p].$$

We write

$$\begin{aligned} \phi_{j,k}(x) &= 2^{j/2} \phi(2^j x - k), & \psi_{j,k}(x) &= 2^{j/2} \psi(2^j x - k), \\ V_j &:= \overline{\text{Span}\{\phi_{j,k} : k \in \mathbb{Z}\}}, & W_j &:= \overline{\text{Span}\{\psi_{j,k} : k \in \mathbb{Z}\}}. \end{aligned}$$

With the above notation, $(V_j)_{j \in \mathbb{Z}}$ is the multiresolution analysis for ϕ , with the conventions

$$V_j \subset V_{j+1}, \quad V_{j+1} = V_j \oplus W_j.$$

where W_j here is the orthogonal complement of V_j in V_{j+1} . For a fixed $J \in \mathbb{N}$ we define the set²

$$B_w := \left\{ \begin{array}{l} \phi_{J,k}, \psi_{j,k} : \text{Supp}(\phi_{J,k}) \cap (-1, 1) \neq \emptyset, \\ \text{Supp}(\psi_{j,k}) \cap (-1, 1) \neq \emptyset, \\ j \in \mathbb{N}, j \geq J, k \in \mathbb{Z} \end{array} \right\}, \quad (30)$$

Let ρ be an ordering of B_w . Notice that since $L^2(\mathbb{R}) = \overline{V_J \oplus \bigoplus_{j=J}^\infty W_j}$ for all $f \in L^2(\mathbb{R})$ with $\text{supp}(f) \subseteq [-1, 1]$ we have

$$f = \sum_{n=1}^\infty c_n \rho(n) \quad \text{for some} \quad (c_n)_{n \in \mathbb{N}} \in \ell^2(\mathbb{N}).$$

Definition 14 (Leveled ordering (standard wavelets)). Define $F_w : B_w \rightarrow \mathbb{R}$ by

$$F_w(f) = \begin{cases} j, & \text{if } f \in W_j \\ -1, & \text{if } f \in V_J \end{cases},$$

and say that any ordering $\tau : \mathbb{N} \rightarrow B_w$ is a ‘leveled ordering’ if it is consistent with F_w .

Notice that $F_w(\psi_{j,k}) = j$. We use the name “leveled” here since requiring an ordering to be leveled means that you can order however you like within the individual wavelet levels themselves, as long as you correctly order the sequence of wavelet levels according to scale.

²‘w’ here stands for ‘wavelet’.

5.4 Boundary Wavelets

We now look at an alternative way of decomposing a function $f \in L^2([-1, 1])$ in terms of a wavelet basis, namely using boundary wavelets [28, Section 7.5.3]. The basis functions all have support contained within $[-1, 1]$, while still spanning $L^2[-1, 1]$. Furthermore, the boundary wavelet basis retains the ability to reconstruct polynomials of order up to $p-1$ from the corresponding standard wavelet basis. We shall not go into great detail here but we will outline the construction; we take, along with a Daubechies wavelet ψ and corresponding scaling function ϕ with $\text{Supp}(\psi) = \text{Supp}(\phi) = [-p+1, p]$, boundary scaling functions and wavelets (using the same notation as in [28] ³)

$$\phi_n^{\text{left}}, \phi_n^{\text{right}}, \psi_n^{\text{left}}, \psi_n^{\text{right}}, \quad n = 0, \dots, p-1.$$

Like in the standard wavelet case we shift and scale these functions,

$$\phi_{j,n}^{\text{left}}(x) = 2^{j/2} \phi_n^{\text{left}}(2^j(x+1)), \quad \phi_{j,n}^{\text{right}}(x) = 2^{j/2} \phi_n^{\text{right}}(2^j(x-1)).$$

We are then able to construct nested spaces, $(V_j^{\text{int}})_{j \geq J}$, for $J \geq \lceil \log_2(p) \rceil$, such that $L^2([-1, 1]) = \overline{\bigoplus_{j=0}^{\infty} V_j^{\text{int}}}$ and $V_{j+1}^{\text{int}} = V_j^{\text{int}} \oplus W_j^{\text{int}}$ by defining

$$V_j^{\text{int}} = \overline{\text{Span} \left\{ \begin{array}{l} \phi_{j,n}^{\text{left}}, \phi_{j,n}^{\text{right}} : n = 0, \dots, p-1 \\ \phi_{j,k} : k \in \mathbb{Z} \text{ s.t. } \text{Supp}(\phi_{j,k}) \subset [-1, 1] \end{array} \right\}},$$

$$W_j^{\text{int}} = \overline{\text{Span} \left\{ \begin{array}{l} \psi_{j,n}^{\text{left}}, \psi_{j,n}^{\text{right}} : n = 0, \dots, p-1 \\ \psi_{j,k} : k \in \mathbb{Z} \text{ s.t. } \text{Supp}(\psi_{j,k}) \subset [-1, 1] \end{array} \right\}}.$$

We then take the spanning elements of V_j^{int} and the spanning elements of W_j^{int} for every $j \geq J$ to form the basis B_{bw} (bw for 'boundary wavelets').

Definition 15 (Leveled ordering (boundary wavelets)). *Define $F_w : B_{\text{bw}} \rightarrow \mathbb{R}$ by the formula*

$$F_{\text{bw}}(f) = \begin{cases} j, & \text{if } f \in W_j^{\text{int}} \\ -1, & \text{if } f \in V_j^{\text{int}} \end{cases}.$$

Then we say that an ordering $\tau : \mathbb{N} \rightarrow B_{\text{bw}}$ of this basis is a 'leveled ordering' if it is consistent with F_{bw} .

Remark 16. *Let $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$. If we require U to be an isometry we must impose the constraint $(2\epsilon)^{-1} \geq 1 + 2^{-J}(p-1)$ otherwise the elements in B_w do not lie in the span of $B_f(\epsilon)$. For convenience we rewrite this as $\epsilon \in I_{J,p}$ where*

$$I_{J,p} := (0, (2 + 2^{-J+1}(p-1))^{-1}].$$

If B_w is replaced by B_{bw} , we only require $\epsilon \leq 1/2$, since every function in B_{bw} has support contained in $[-1, 1]$. For the rest of this section, we shall assume these constraints on ϵ hold.

6 1D Fourier-Wavelet Case

Let $U = [(B_f(\epsilon), \rho), (B_2, \tau)]$ with either $B_2 = B_w$ or B_{bw} . The key observation for handling the entries of U are

$$\begin{aligned} U_{m,n} = \langle \tau(n), \rho(m) \rangle &= \int_{\mathbb{R}} \sqrt{\epsilon} \exp(-2\pi i \epsilon x \cdot \lambda \circ \rho(m)) \cdot \tau(n)(x) \, dx \\ &= \sqrt{\epsilon} \mathcal{F}\tau(n)(\epsilon \cdot \lambda \circ \rho(m)), \end{aligned} \tag{31}$$

³We use $[-1, 1]$ instead of $[0, 1]$ as our reconstruction interval here, but everything else is the same.

where \mathcal{F} denotes the 1D Fourier Transform. We also observe that

$$\begin{aligned}\mathcal{F}\phi_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\phi(2^{-j}\omega), & \mathcal{F}\psi_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\psi(2^{-j}\omega), \\ \mathcal{F}\psi_{j,n}^{\text{left}}(\omega) &= 2^{-j/2} e^{2\pi i} \mathcal{F}\psi_n^{\text{left}}(2^{-j}\omega), & \mathcal{F}\psi_{j,n}^{\text{right}}(\omega) &= 2^{-j/2} e^{-2\pi i} \mathcal{F}\psi_n^{\text{right}}(2^{-j}\omega).\end{aligned}\tag{32}$$

Outline of Argument: Suppose ρ is a frequency ordering and τ a leveled ordering. When we bound $\mu(\pi_N U)$ and $\mu(U\pi_N)$ we rely heavily on the formula

$$|U_{m,n}|^2 = \epsilon 2^{-j(n)} |\mathcal{F}\psi(2^{-j(n)} \epsilon \cdot \lambda \circ \rho(m))|^2, \tag{33}$$

where $j(n)$ denotes what wavelet level we are on and we also ignore scaling function terms.

If we want to bound $\mu(U\pi_N)$, i.e. looking at a single wavelet function and maximising over all frequencies, then the $|\mathcal{F}\psi(\cdot)|^2$ term varies little as $j(N)$ gets large and we just have $\mu(\pi_N U) = \epsilon \Theta(2^{-j(N)})$. There are roughly $2^{j(N)}$ functions in the wavelet basis for each value of $j(N)$, implying that $2^{-j(N)} = \Theta(N^{-1})$ and therefore $\mu(U\pi_N) = \epsilon \Theta(N^{-1})$.

When upper bounding $\mu(\pi_N U)$, one uses the decay property $|\psi(\omega)| \leq K \cdot |\omega|^{-1/2}$ to remove the dependence on $j(n)$ in (33), leaving us with $\mu(\pi_N U) = \mathcal{O}(|\lambda \circ \rho(m)|^{-1})$. Observing $|\lambda \circ \rho(m)| = \Theta(m)$ gives us $\mu(\pi_N U) = \mathcal{O}(N^{-1})$. For the lower bound we look near the diagonal of the matrix $n = m$. Recalling $2^{j(m)}, |\lambda \circ \rho(m)| = \Theta(m)$ and noticing that $2^{-j(m)} \epsilon |\lambda \circ \rho(m)|$ is roughly constant, this suggests that $|\mathcal{F}\psi(2^{-j(m)} \epsilon |\lambda \circ \rho(m)|)|^2$ is nearly constant and we're left with $\mu(\pi_N U) \gtrsim 2^{-j(N)} = \Theta(N^{-1})$.

We now come to our first concrete example of bounding line coherences

Theorem 17. *Let $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$ where τ is a leveled ordering of a standard wavelet basis. Then there are constants $C_1, C_2 > 0$, dependent on the choice of wavelet, such that for all $\epsilon \in I_{J,p}$ and $N \in \mathbb{N}$, we have*

$$\frac{\epsilon \cdot C_1}{N} \leq \mu(U\pi_N) \leq \frac{\epsilon \cdot C_2}{N}. \tag{34}$$

Furthermore, suppose instead $U = [(B_f(\epsilon), \rho), (B_{bw}, \tau)]$ where τ is a leveled ordering of a boundary wavelet basis. Then there are constants $C_1, C_2 > 0$, dependent on the choice of wavelet, such that for all $\epsilon \in (0, 1/2]$ and $N \in \mathbb{N}$, (34) holds.

Proof. By equation (31) we know that (since $\lambda \circ \rho : \mathbb{N} \rightarrow \mathbb{Z}$ is bijective)

$$\mu(\pi_N U) = \sup_{m \in \mathbb{N}} \epsilon |\mathcal{F}\tau(N)(\epsilon \cdot \lambda \circ \rho(m))|^2 = \sup_{m \in \mathbb{Z}} \epsilon |\mathcal{F}\tau(N)(\epsilon m)|^2.$$

Case 1 (Standard wavelets): In this case we define $j(N) := F_w(\tau(N))$ and let $a := 2p - 1 \in \mathbb{N}$ denote the length of the support of the scaling function ϕ corresponding to B_w . Notice that for a leveled ordering of B_w , the functions belonging to V_J come first, and there are of $2^{J+1} + a - 1$ of these functions. Therefore, for $N \leq 2^{J+1} + a - 1$ we have that, by (32),

$$\mu(U\pi_N) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-J} |\mathcal{F}\phi(2^{-J} \epsilon m)|^2. \tag{35}$$

Furthermore, for the wavelet terms in B_w , which correspond to $N \geq 2^{J+1} + a$, we have that, by (32),

$$\mu(U\pi_N) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-j(N)} |\mathcal{F}\psi(2^{-j(N)} \epsilon m)|^2. \tag{36}$$

Since the wavelet is compactly supported and in $L^2(\mathbb{R})$ it is in $L^1(\mathbb{R})$ and so its Fourier transform is continuous. Notice that by continuity and the Riemann-Lebesgue Lemma, we see that $\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)| = |\mathcal{F}\psi(\hat{\omega})|$ for some $\hat{\omega} \in \mathbb{R}$. Therefore, since $j(N) \rightarrow \infty$ as $N \rightarrow \infty$ because the ordering τ is leveled, we find that

$$\frac{\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(\epsilon 2^{-j(N)} m)|^2}{\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2} \longrightarrow 1 \quad \text{as } N \rightarrow \infty. \tag{37}$$

Furthermore, this convergence is uniform in $\epsilon \in I_{J,p}$ as $N \rightarrow \infty$. We are therefore left with handling the $2^{-j(N)}$ term, which means estimating $j(N)$ as $N \rightarrow \infty$.

Notice that for each value of $j(N) \geq J$ there are $2^{j(N)+1} + a - 1$ functions in our wavelet basis with this value of $j(N)$. For simplicity we shall use the simple bounds $2^{j(N)+1} \leq 2^{j(N)+1} + a - 1 \leq 2^{j(N)+a}$. Now for every $N \in \mathbb{N}$ with $j(N) > J$, we must have had all the terms of the form $f \in B_w, F_w(f) = j(N) - 1$ come before N in the leveled ordering and there are at least $2^{j(N)}$ of these terms. If $j(N) = J$ we instead have $N > 2^{J+1} + a - 1 > 2^J$. Likewise for every $N \in \mathbb{N}$ with $j(N) \geq J$ there can be no more than $\sum_{i=J}^{j(N)} 2^{(i+a)} + 2^J + a - 1 \leq 2^{j(N)+a+2}$ terms that came before N . Therefore, for $j(N) \geq J$, we have the inequality

$$2^{j(N)} \leq N \leq 2^{j(N)+a+2}. \quad (38)$$

From here we will tackle the upper and lower bounds of (34) separately:

Upper Bound: We will show that $\mu(U\pi_N) \leq \frac{\epsilon C_2}{N}$. Notice from (38) we have the upper bound $2^{-j(N)} \leq 2^{a+2} N^{-1}$ for $j(N) \geq J$ and therefore for these terms we can bound (36) by

$$\epsilon \sup_{m \in \mathbb{Z}} 2^{-j(N)} |\mathcal{F}\psi(2^{-j(N)} \epsilon m)|^2 \leq \epsilon \frac{2^{a+2}}{N} \cdot \sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2,$$

For the $j(N) = -1$ terms (i.e. $N \leq 2^{J+1} + a - 1$) we also have the simple bound

$$\epsilon \cdot 2^{-J} \sup_{\omega \in \mathbb{R}} |\mathcal{F}\phi(\omega)|^2 \leq \epsilon 2^{-J} \frac{2^{J+1} + a - 1}{N} \cdot \sup_{\omega \in \mathbb{R}} |\mathcal{F}\phi(\omega)|^2,$$

and so the upper bound is complete.

Lower Bound: First notice that from (38) that we have the lower bound $2^{-j(N)} \geq N^{-1}$. Next notice that from (37) there is an $N' \in \mathbb{N}$ independent of $\epsilon \in (0, 1/2p]$ such that for all $N \geq N'$ we have

$$\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(\epsilon 2^{-j(N)} m)|^2 \geq \frac{1}{2} \sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2.$$

Consequently for $N \geq N'$ we have the lower bound

$$\mu(\pi_N U) \geq \epsilon 2^{-j(N)} \cdot \frac{\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2}{2} \geq \frac{\epsilon}{2N} \cdot \sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2.$$

Therefore, in order to prove the lower bound, we need only show there exists a constant $C > 0$ such that every $N < N'$ we have $\mu(\pi_N U) \geq \epsilon \cdot C$ uniformly in $\epsilon \in I_{J,p}$. This will be satisfied if we can show that for every $j \geq J$ fixed there exists a constant $C > 0$ such that for all $\epsilon \in I_{J,p}$

$$\sup_{m \in \mathbb{Z}} |\mathcal{F}\phi(2^{-j} \epsilon m)|^2, \sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(2^{-j} \epsilon m)|^2 \geq C.$$

We will deal with latter term since the scaling function term is handled similarly. We know that for every $\epsilon \in I_{J,p}$ fixed, $\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(2^{-j} \epsilon m)|^2 > 0$ since if it were not the case we would find that $\langle \chi_m, \psi_{j,0} \rangle = 0$ for every m , contradicting the χ_m forming a basis of $L^2([-2\epsilon)^{-1}, (2\epsilon)^{-1}])$. Next notice that by the Riemann-Lebesgue Lemma and continuity of the Fourier transform of ψ , this supremum is a continuous function of ϵ and that

$$\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(2^{-j} \epsilon m)|^2 \rightarrow \sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2 > 0 \quad \text{as } \epsilon \rightarrow 0.$$

Consequently we deduce the supremum attains its lower bound as a function of ϵ on $I_{J,p}$ and we are done.

Case 2 (Boundary wavelets): The method of proof is the same except that we have additional $\psi^{\text{left}}, \phi^{\text{left}}, \psi^{\text{right}}, \phi^{\text{right}}$ terms to deal with. We also have slightly different behaviour of $2^{j(N)}$, i.e. for $N > 2^{J+1}$,

$$2^{j(N)} \leq N \leq 2^{j(N)+2}. \quad (39)$$

This follows from observing that for each value of $j(N)$ there are $2^{j(N)+1}$ functions in the wavelet basis, and that we are using a leveled ordering. The details are omitted for the sake of brevity. \square

Next we need the following condition on our scaling function / wavelet; there exists a constant $K > 0$ s.t. $\forall \omega \in \mathbb{R} \setminus \{0\}$,

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \quad (40)$$

This condition holds for all Daubechies wavelets (see the proof of Proposition 4.7 in [13]), in fact it even holds if we change the power of ω from $1/2$ to 1 .

Lemma 18. *Let ϕ be a Daubechies scaling function, with corresponding mother wavelet ψ . Then, along with (40), we also have*

$$|\mathcal{F}\psi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \quad (41)$$

Furthermore in the case of boundary wavelets we also have for some constant $K > 0$ and $\omega \in \mathbb{R} \setminus \{0\}$

$$|\mathcal{F}\phi_n^{\text{left}}(\omega)|, |\mathcal{F}\phi_n^{\text{right}}(\omega)|, |\mathcal{F}\psi_n^{\text{left}}(\omega)|, |\mathcal{F}\psi_n^{\text{right}}(\omega)| \leq \frac{K}{|\omega|^{1/2}}, \quad (42)$$

along with (40) and (41). In fact (41) and (42) hold with the powers of $1/2$ replaced by 1 .

Proof. We notice that if (40) holds then we can use the equation (from (2.14) in [21])

$$\mathcal{F}\psi(2\omega) = \exp(2i\pi\omega) \cdot \nu(2\omega) \cdot m_0(\omega + 1/2) \cdot \mathcal{F}\phi(\omega), \quad (43)$$

where m_0 is the Fourier transform of the low pass filter of the scaling function ϕ and ν is function whose modulus is always 1⁴. Taking the modulus of this equation gives $|\mathcal{F}\psi(2\omega)| = |m_0(\omega + 1/2)| \cdot |\mathcal{F}\phi(\omega)|$. Therefore using this along with $|m_0(\omega)| \leq 1$, $\forall \omega \in \mathbb{R}$ (from (2.5) in [21]) we can show that (40) also holds with ϕ replaced by ψ .

We now turn to the boundary wavelet estimates. We may assume $p \geq 2$ since in the Haar case boundary wavelets are redundant. First we note that the property of having a decay estimate of the form 40 is closed under finite linear combinations. Next observe that if we prove an estimate of the form 40 for the functions (see page 71 of [12])

$$\tilde{\phi}^k(x) = \sum_{n=k}^{2p-2} \binom{n}{k} \phi(x + n - p + 1) \cdot \mathbb{1}_{[0, \infty)}, \quad k = 0, \dots, p-1,$$

then we also have the same decay (with a different constant) for the functions ϕ_k^{left} and ψ_k^{left} since they are finite linear combinations of these functions. A similar argument will work for the right boundary wavelets. Let us consider an arbitrary term from the sum

$$T_n(x) := \phi(x + n - p + 1) \cdot \mathbb{1}_{[0, \infty)} = \phi(x + n - p + 1) \cdot \mathbb{1}_{[0, 2p-1]}.$$

Now since we have expressed T_n as a product of two L^2 functions we can apply the convolution rule on its Fourier Transform to deduce $\mathcal{F}T_n(\omega) = (\mathcal{F}\phi_{0, -n+p-1} * \mathcal{F}\mathbb{1}_{[0, 2p-1]})(\omega)$. Now we make two observations:

1. $|\mathcal{F}\mathbb{1}_{[0, 2p-1]}(\omega)| = |(\exp(-2\pi i(2p-1)\omega) - 1) \cdot (2\pi i\omega)^{-1}| \leq C_1 \cdot (|\omega| + 1)^{-1}$ for some constant $C_1 > 0$.

2. Excluding the Haar wavelet, for every Daubechies wavelet there exists constants $\alpha, C_2 > 0$ such that $|\mathcal{F}\phi(\omega)| \leq C_2 \cdot (|\omega| + 1)^{-1-\alpha}$ (see the proof of Proposition 4.7 in [13]).

We now claim that if two functions f, g satisfy

$$|f(\omega)| \leq C_1 \cdot (|\omega| + 1)^{-1}, \quad |g(\omega)| \leq C_2 \cdot (|\omega| + 1)^{-1-\alpha}, \quad \forall \omega \in \mathbb{R}.$$

⁴The equation here is not identical to that of the reference because of our choice of definition of the Fourier transform.

for some constants $\alpha, C_1, C_2 > 0$ then $|f * g(\omega)| \leq C_3 \cdot |\omega|^{-1}$ which will prove the lemma. To see this notice that (without loss of generality $\omega > 0$)

$$\begin{aligned} |f * g(\omega)| \cdot |\omega| &\leq C_1 C_2 \int_{\mathbb{R}} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \\ &\leq C_1 C_2 \left(\int_{-\infty}^{\omega/2} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \right. \\ &\quad \left. + \int_{\omega/2}^{+\infty} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \right), \end{aligned} \quad (44)$$

and notice that we would have shown the claim if we can bound the RHS uniformly in ω . By noting $|\omega - u| + 1 \geq |u| + 1$, $|\omega - u| \geq |\omega/2|$ for $u \in (-\infty, \omega/2]$ we see that the first integral is bounded above by

$$\begin{aligned} \int_{-\infty}^{\omega/2} \frac{|\omega|}{(|u| + 1)^{1+\alpha/2}(|\omega - u| + 1)^{1+\alpha/2}} du &\leq \int_{-\infty}^{\omega/2} \frac{|\omega|}{(|u| + 1)^{1+\alpha/2}(|\omega/2| + 1)^{1+\alpha/2}} du \\ &\leq \int_{\mathbb{R}} \frac{2^{1+\alpha/2}}{(|u| + 1)^{1+\alpha/2}} du = \text{constant} < \infty. \end{aligned}$$

To bound the last integral in (44) we simply use $|\omega|(|u| + 1)^{-1} \leq 2$ for $u \in [\omega/2, \infty)$ to give us a similar uniform upper bound, completing the proof of the claim. \square

For our second incoherence result we will need a technical lemma.

Lemma 19. *For any compactly supported wavelet ψ with scaling function $\phi \in L^1(\mathbb{R})$ there exists an $N \in \mathbb{N}$ such that for all $q \geq N$, ($q \in \mathbb{N}$) we have*

$$L_q := \inf_{\omega \in [2^{-(q+1)}, 2^{-q}]} |\mathcal{F}\psi(\omega)| > 0.$$

Proof. We recall from equation (43) that

$$|\mathcal{F}\psi(2\omega)| = |m_0(\omega + 1/2)| \cdot |\mathcal{F}\phi(\omega)|. \quad (45)$$

Furthermore, we also know that $|\mathcal{F}\phi(0)| = 1$ and $m_0(1/2) = 0$ [21]⁵. However, since ϕ is compactly supported, m_0 is a non-zero trigonometric polynomial and so it follows that this zero at $1/2$ is isolated. Therefore, since $\mathcal{F}\phi$ is continuous, we deduce that (45) is nonzero when $\omega > 0$ is sufficiently small. \square

Now we cover the second half of our line-coherence bounds

Theorem 20. *Let $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$ where ρ is a frequency ordering of the Fourier basis. Then there is a constant $C_1 > 0$ such that for all $\epsilon \in I_{J,p}$ and $N \in \mathbb{N}$, we have the upper bound*

$$\mu(\pi_N U) \leq \frac{C_1}{N}.$$

Furthermore, there is a constant $C_2 > 0$ such that for all $\epsilon \in I_{J,p}$ and $N \geq 1 + 2^{J+1}\epsilon^{-1}$ we have the lower bound

$$\mu(\pi_N U) \geq \frac{C_2}{N}.$$

Finally, if we replace B_w by B_{bw} in the above setup, the same conclusions also hold with the constraint $\epsilon \in I_{J,p}$ replaced by $\epsilon \in (0, 1/2]$.

⁵See Section 2 Theorem 1.7 and Equation (3.1) in the reference.

Proof. Upper Bound: Since ρ is a frequency ordering if $m = 1$ then $\lambda \circ \rho(m) = 0$ and since $|\mathcal{F}\phi(0)| = 1$, $\mathcal{F}\psi(0) = 0$ (see (45) and the line below it), in the case of standard wavelets we have $\mu(\pi_1 U) = \epsilon 2^{-J}$. In the case of boundary wavelets we have the estimate

$$\mu(\pi_1 U) \leq \epsilon \cdot 2^{-J} \cdot \max(1, |\psi^{\text{left}}(0)|, |\psi^{\text{right}}(0)|, |\phi^{\text{left}}(0)|, |\phi^{\text{right}}(0)|)^2,$$

Next let $m \geq 1$. For standard wavelets we observe that the estimate (40) is strong enough to bound the finitely many $\phi_{J,k}$ terms as required since

$$|\langle \phi_k, \rho(m) \rangle|^2 = \epsilon 2^{-J} |\mathcal{F}\phi(\epsilon 2^{-J} \cdot \lambda \circ \rho(m))|^2 \leq \frac{\epsilon 2^{-J} \cdot K^2}{|\epsilon 2^{-J} \cdot \lambda \circ \rho(m)|} \leq \frac{2K^2}{m-1},$$

where we used that ρ is a frequency ordering in the last step (for boundary wavelets the same holds for the finitely many V_J^{int} terms). Therefore we are left with the terms involving the shifts and dilations of ψ (and for boundary wavelets the $\psi_k^{\text{left}}, \psi_k^{\text{right}}$ terms as well). This is also a straightforward consequence of (40) since we have

$$\begin{aligned} |\langle \psi_{j,k}, \rho(m) \rangle|^2 &= \epsilon 2^{-j} |\mathcal{F}\psi(\epsilon 2^{-j} \cdot \lambda \circ \rho(m))|^2 \\ &\leq \epsilon 2^{-j} \cdot \frac{2^j K^2}{\epsilon \cdot |\lambda \circ \rho(m)|} \leq \frac{K^2}{|\lambda \circ \rho(m)|} \leq \frac{2K^2}{m-1}, \end{aligned}$$

and for boundary wavelets we can tackle the $\psi_k^{\text{left}}, \psi_k^{\text{right}}$ terms in the same way. This gives the global bound for $m \geq 2$ (uniform in n and ϵ)

$$|\langle \tau(n), \rho(m) \rangle|^2 \leq \frac{2K^2}{m-1} \leq \frac{4K^2}{m}.$$

Combining this with our bound on $\mu(\pi_1 U)$ (we just bound ϵ by 1) we obtain the required upper bound.

Lower Bound: For standard wavelets, given $m \in \mathbb{N}$, $m \neq 1$, find $n \in \mathbb{N}$ such that $\tau(n) = \psi_{j,0}$ with $j = \lceil \log_2(\epsilon |\lambda \circ \rho(m)|) \rceil + q$, where $q \in \mathbb{N}$ is arbitrary but sufficiently large so that $j \geq J$. Notice that this means that $\epsilon 2^{-j} |\lambda \circ \rho(m)| \in (2^{-q-1}, 2^{-q}]$. Therefore, recalling the definition of L_q in Lemma 19, we see that we have

$$\begin{aligned} |\langle \tau(n), \rho(m) \rangle|^2 &= \epsilon 2^{-j} |\mathcal{F}\psi(2^{-j} \epsilon \lambda \circ \rho(m))|^2 \\ &\geq \epsilon 2^{-\lceil \log_2(\epsilon |\lambda \circ \rho(m)|) \rceil - q} |\mathcal{F}\psi(\epsilon \cdot 2^{-\lceil \log_2(\epsilon |\lambda \circ \rho(m)|) \rceil - q} \cdot \lambda \circ \rho(m))|^2 \\ &\geq \frac{L_q^2 \cdot 2^{-q}}{2|\lambda \circ \rho(m)|} \geq \frac{L_q^2 \cdot 2^{-q}}{m}. \end{aligned}$$

We used $m \neq 1$ in the last step and the fact that the ordering ρ is standard. Recall that by Lemma 19 there exists a $q \in \mathbb{N}$ such that $L_q > 0$. We choose the same such q for all $\epsilon \in I_{J,p}$. To ensure that $j = \lceil \log_2(\epsilon |\lambda \circ \rho(m)|) \rceil + q$ satisfies $j \geq J$ we must therefore impose the constraint that m is sufficiently large. $j \geq J$ is satisfied if

$$J \leq \log_2(\epsilon |\lambda \circ \rho(m)|) \quad \Leftrightarrow \quad m \geq 1 + 2^{J+1} \epsilon^{-1}.$$

When using boundary wavelets the argument for the lower bound is identical. □

Remark 21. The condition $N \geq 1 + 2^{J+1} \epsilon^{-1}$ cannot be replaced by $N \in \mathbb{N}$ for the lower bound since, in the case of standard wavelets, for every fixed $N \in \mathbb{N}$ we have

$$\mu(\pi_N U) \leq \epsilon \cdot \max \left(\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2, \sup_{\omega \in \mathbb{R}} |\mathcal{F}\phi(\omega)|^2 \right) = \mathcal{O}(\epsilon).$$

Summing up Theorems 20 and 17 without the ϵ dependence we have

Corollary 22. Let $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$ where ρ is a frequency ordering, τ is a leveled ordering and $\epsilon \in I_{J,p}$. Then

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1}). \quad (46)$$

If B_w is replaced by B_{bw} and $\epsilon \in I_{J,p}$ by $\epsilon \in (0, 1/2]$ then (46) also holds.

7 1D Fourier-Polynomial Case

Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is frequency ordering and τ the natural ordering on the Legendre polynomials. Before we start formally covering the line coherence estimates for these two bases we shall first need to prove a preliminary result.

Outline of Argument: The key fact we use in this section is that the entries of U are directly related to Bessel functions (see (51):

$$\begin{aligned} |U_{m,n}| &= 2\sqrt{\epsilon(n-1/2)} \cdot |j_{n-1}(2\pi\epsilon\lambda \circ \rho(m))| \\ &= \frac{\sqrt{n-1/2}}{\sqrt{\lambda \circ \rho(m)}} \cdot |J_{n-1/2}(2\pi\epsilon\lambda \circ \rho(m))|, \quad (m \neq 1), \end{aligned} \quad (47)$$

where J_n is a Bessel function of the first kind and j_n is a spherical Bessel function of the first kind. From here we rely heavily upon various asymptotic results regarding J_n, j_n to produce the appropriate bounds. For $\mu(U\pi_N)$ we maximise over m in (2) suggesting that $\mu(U\pi_N) = \Theta(\epsilon(N-1/2) \cdot \sup_{\mathbb{R}} |j_{N-1}|^2)$. Since $\sup_{\mathbb{R}} |j_{N-1}|^2 = \Theta(N^{-5/6})$ this case is then complete. The case of bounding $\mu(\pi_N U)$ is more involved and requires breaking down suprema into the cases $m > n$ and $m \leq n$ for the upper bound and then looking near the diagonal for the lower bound.

Lemma 23. *Let J_n denote the n th Bessel function of the first kind and let $j'_{n,k}$ denote the k th non-negative root of J'_n . Furthermore, Let j_n denote the n th spherical Bessel function of the first kind and let $a'_{n,k}$ denote the k th non-negative root of j'_n . Then if $n \geq 1$ we have*

$$\sup_{x \in \mathbb{R}} |J_n(x)| = |J_n(j'_{n,1})|, \quad \sup_{x \in \mathbb{R}} |j_n(x)| = |j_n(a'_{n,1})|.$$

Proof. The result for J_n follows from the arguments given in [32, Section 15]. Instead of repeating them here again, we instead adapt the same approach to deduce the Lemma for j_n . We will be using two facts about j_n . First, we have the power series expansion [1, Eqn. (10.1.2)]

$$j_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m 2^{n+1} (n+m+1)! x^{n+2m}}{m! (2(n+m+1))!}. \quad (48)$$

Second, we shall use the fact that j_n is a solution to the following differential equation [1, Eqn. (10.1.1)]

$$x^2 j_n(x)'' + 2x j_n'(x) + (x^2 - n(n+1)) j_n(x) = 0. \quad (49)$$

We first observe that by (48), $|j_n(-x)| = |j_n(x)|$, $\forall x \in \mathbb{R}$ and so we need only consider $\sup_{x \in [0, +\infty)} |j_n(x)|$. (49) can be rephrased as

$$(x^2 j_n'(x))' = (n(n+1) - x^2) j_n(x).$$

Therefore, noting that by (48), $j_n(x) > 0$ for $x > 0$ sufficiently small, we deduce that $x^2 j_n'(x)$ is positive for $x \in (0, n(n+1)]$ and hence so is $j_n'(x)$. This tells us that $a'_{n,k} > n(n+1)$ for all $k \in \mathbb{N}$.

Now consider the function

$$\Lambda_n(x) := j_n^2(x) + \frac{x^2 j_n'^2(x)}{x^2 - n(n+1)}, \quad x \in (n(n+1), +\infty).$$

Observe that $\Lambda_n(a'_{n,k}) = j_n^2(a'_{n,k})$ for all $n, k \in \mathbb{N}$. Moreover the derivative is always negative:

$$\begin{aligned} \Lambda_n'(x) &= 2j_n'(x)j_n(x) + \frac{2x j_n'^2(x) + 2x^2 j_n'(x)j_n''(x)}{x^2 - n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2 - n(n+1))^2} \\ &= \frac{2j_n'(x)(j_n(x)(x^2 - n(n+1)) + x j_n'(x) + x^2 j_n''(x))}{x^2 - n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2 - n(n+1))^2} \\ &= -\frac{2x j_n'^2(x)}{x^2 - n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2 - n(n+1))^2} < 0. \quad (\text{using (49)}) \end{aligned} \quad (50)$$

This tells that $|j_n(a'_{n,1})| > |j_n(a'_{n,2})| > |j_n(a'_{n,3})| \dots$. To finish the proof we notice that by (48), $j_n(0) = 0$ for $n \geq 1$ and furthermore, by [1, Eqn. (10.1.14)],

$$j_n(x) = \frac{(-i)^n}{2} \int_{-1}^1 e^{ixt} R_N(t) dt,$$

and therefore $j_n(x) \rightarrow 0$ as $x \rightarrow +\infty$ by the Riemann-Lebesgue Lemma. We therefore know that the maxima of $|j_n(x)|$ on $[0, +\infty)$ must be attained at its first stationary point. \square

Theorem 24. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where τ is the natural ordering of the polynomial basis. Then there are constants $C_1, C_2 > 0$ such that for all $\epsilon \in (0, 0.45]$ and $N \in \mathbb{N}$,*

$$\frac{\epsilon \cdot C_1}{N^{2/3}} \leq \mu(U\pi_N) \leq \frac{\epsilon \cdot C_2}{N^{2/3}}.$$

Proof. Upper Bound: First notice that

$$\begin{aligned} U_{m,n} &= \langle \rho(m), \tilde{p}_n \rangle_{L^2([-1,1])} \\ &= \sqrt{\epsilon} \cdot \sqrt{n-1/2} \int_{-1}^1 e^{2\pi i \lambda \circ \rho(m) \epsilon t} R_N(t) dt \\ &= i^{n-1} 2\sqrt{\epsilon(n-1/2)} \cdot j_{n-1}(2\pi \epsilon \lambda \circ \rho(m)) \\ &= i^{n-1} \frac{\sqrt{n-1/2}}{\sqrt{\lambda \circ \rho(m)}} \cdot J_{n-1/2}(2\pi \epsilon \lambda \circ \rho(m)), \quad (m \neq 1) \end{aligned} \tag{51}$$

where on the third line we have used [1, Eqn. (10.1.14)] and on the fourth line we have used the following formula connecting the spherical Bessel function to the standard Bessel function:

$$j_n(z) = \sqrt{\frac{\pi}{2z}} J_{n+1/2}(z). \tag{52}$$

Therefore, we find

$$\mu(U\pi_N) \leq 4\epsilon(N-1/2) \sup_{t \in \mathbb{R}} j_{N-1}^2(t). \tag{53}$$

We therefore need to estimate $\sup_{t \in \mathbb{R}} |j_n(t)|$. By Lemma 23, we know that $\sup_{t \in \mathbb{R}} |j_n(t)| = |j_n(a'_{n,1})|$ for $n \geq 1$, where $a'_{n,1}$ denotes the first positive root of j'_n .

Thus, we only need to have estimates for $|j_n(a'_{n,1})|$. But we also know [1, Eqn. (10.1.61)], that the following asymptotic expansion holds

$$j_n(a'_{n,1}) \sim \gamma(n+1/2)^{-5/6} + \mathcal{O}((n+1/2)^{-3/2}), \tag{54}$$

for some positive constant $1/2 < \gamma < 1$. Therefore we know there exists $N' \in \mathbb{N}$ such that for all $N > N'$ we have

$$\sup_{x \in \mathbb{R}} |j_N(x)| \leq (N+1/2)^{-5/6}.$$

Applying this bound to (53) we get the upper bound

$$\begin{aligned} \mu(U\pi_N) &\leq 4\epsilon(N-1/2) \cdot (N-1/2)^{-5/3} \\ &\leq \frac{4\epsilon}{(N-1/2)^{2/3}} \leq \frac{8\epsilon}{N^{2/3}}. \end{aligned}$$

Therefore the upper bound is complete for the case $N > N'$ (and notice that N' is independent of ϵ). However since $\sup_{x \in \mathbb{R}} |j_N(x)| < \infty$ for every N we can use (53) to cover the case $N \leq N'$, completing the upper bound.

Lower Bound: We focus on the following equation taken from (51)

$$|U_{m,n}| = \frac{\sqrt{n-1/2}}{\sqrt{|\lambda \circ \rho(m)|}} \cdot |J_{n-1/2}(2\pi\epsilon\lambda \circ \rho(m))|, \quad (m \neq 1). \quad (55)$$

Let $j'_{\nu,1}$ denote the first positive zero of J'_ν . From [1, Eqns. (9.5.16), (9.5.20)], we have the asymptotic estimates

$$j'_{\nu,1} \sim \nu + \zeta\nu^{1/3} + \mathcal{O}(\nu^{-1/3}), \quad (56)$$

$$J(j'_{\nu,1}) \sim \kappa \cdot \nu^{-1/3} + \mathcal{O}(\nu^{-1}), \quad (57)$$

where $\kappa, \zeta > 0$ are some constants. Next let k_m denote the nearest integer multiple of $2\pi\epsilon$ to $j'_{n-1/2,1}$, which means that $|k_n - j'_{n-1/2,1}| \leq \pi\epsilon$. We shall first prove a lower bound for $|J_{n-1/2}(k_n)|$. Before we do so, we need the following two results:

1. $2J'_\nu(x) = J_{\nu-1}(x) - J_{\nu+1}(x)$, [32, p. 45],
2. $\sup_{x \in \mathbb{R}} |J_\nu(x)| = |J_\nu(j'_{\nu,1})|$, using Lemma 23.

These two results can be combined to give us $\sup_{x \in \mathbb{R}} |J''_\nu(x)| \leq |J_\nu(j'_{\nu,1})|$, which we will use in (58) below.

By the triangle inequality $|J_{n-1/2}(k_n)| \geq |J_{n-1/2}(j'_{n-1/2,1})| - |J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})|$ and we bound the latter term by using integrals⁶:

$$\begin{aligned} |J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})| &= \left| \int_{j'_{n-1/2,1}}^{k_n} J'_{n-1/2,1}(t) dt \right| \\ &= \left| \int_{j'_{n-1/2,1}}^{k_n} \int_{j'_{n-1/2,1}}^t J''_{n-1/2,1}(u) du dt \right| \\ &\leq \left| \int_{j'_{n-1/2,1}}^{k_n} \int_{j'_{n-1/2,1}}^t |J_{n-1/2}(j'_{n-1/2,1})| du dt \right| \\ &\leq \frac{|j'_{n-1/2,1} - k_n|^2}{2} \cdot |J_{n-1/2}(j'_{n-1/2,1})| \\ &\leq \frac{(\pi\epsilon)^2}{2} \cdot |J_{n-1/2}(j'_{n-1/2,1})|. \end{aligned} \quad (58)$$

Notice that there is a constant $1 > d > 0$ such that for all $\epsilon \in (0, 0.45]$ we have $(\pi\epsilon)^2/2 \leq d$ and therefore (58) becomes

$$|J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})| \leq d \cdot |J_{n-1/2}(j'_{n-1/2,1})|,$$

and therefore we deduce

$$\begin{aligned} |J_{n-1/2}(k_n)| &\geq |J_{n-1/2}(j'_{n-1/2,1})| - |J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})| \\ &\geq (1-d) \cdot |J_{n-1/2}(j'_{n-1/2,1})|. \end{aligned}$$

Combining this inequality with (56), (57) gives us the following bound:

$$\begin{aligned} \frac{\sqrt{n-1/2}}{\sqrt{k_n}} \cdot |J_{n-1/2}(k_n)| &\geq \frac{\sqrt{j'_{n-1/2,1}}}{\sqrt{k_n}} \cdot \frac{\sqrt{n-1/2}}{\sqrt{j'_{n-1/2,1}}} \cdot (1-d) |J_{n-1/2}(j'_{n-1/2,1})| \\ &= \frac{\sqrt{j'_{n-1/2,1}}}{\sqrt{k_n}} \cdot \frac{\sqrt{n-1/2}}{\sqrt{n-1/2 + \mathcal{O}(n^{1/3})}} \cdot (1-d)(\kappa(n-1/2)^{-1/3} + \mathcal{O}(n^{-1})). \end{aligned}$$

⁶The use of the second integral is valid since $J'_{n-1/2,1}(t) = J'_{n-1/2,1}(t) - J'_{n-1/2,1}(j'_{n-1/2,1})$ by the definition of $j'_{n-1/2,1}$.

The first two fractions on the last line converge to 1 as $n \rightarrow \infty$ and therefore we deduce that there is an $M \in \mathbb{N}$ and a constant $C > 0$ (independent of ϵ) such that for all $n \geq M$ we have

$$\frac{\sqrt{n-1/2}}{\sqrt{k_n}} \cdot |J_{n-1/2}(k_n)| \geq Cn^{-1/3}. \quad (59)$$

Therefore, given $n \geq M$, let $m(n) \in \mathbb{N}$ be such that $2\pi\epsilon\lambda \circ \rho(m) = k_n$. Then by (59) and (55) we have

$$|U_{m(n),n}| = \sqrt{2\pi\epsilon} \cdot \frac{\sqrt{n-1/2}}{\sqrt{k_n}} \cdot |J_{n-1/2}(k_n)| \geq \sqrt{2\pi\epsilon} \cdot Cn^{-1/3}.$$

Consequently we deduce that $\mu(U\pi_N) \geq 2\pi\epsilon \cdot C^2 N^{-2/3}$ for $N \geq M$.

For $N \leq M$ we observe that from (51)

$$\mu(U\pi_N) = 4\epsilon(N-1/2) \sup_{m \in \mathbb{Z}} |j_{N-1}(2\pi\epsilon m)|^2.$$

As before we observe that since $j_{N-1}(x) \rightarrow 0$ as $x \rightarrow \infty$, the supremum $\sup_{m \in \mathbb{Z}} |j_{N-1}(2\pi\epsilon m)|$ is a continuous function of ϵ and moreover the supremum converges to $\sup_{x \in \mathbb{R}} |j_{N-1}(x)| > 0$ as $\epsilon \rightarrow 0$. Therefore by compactness of $[0, 0.45]$, we know there is a constant $D_N > 0$ such that for all $\epsilon \in (0, 0.45]$ we have

$$\mu(U\pi_N) = 4\epsilon(N-1/2) \cdot D_N^2.$$

This combined with the result $\mu(U\pi_N) \geq 2\pi\epsilon \cdot C^2 N^{-2/3}$ for $N \geq M$ gives us the required lower bound. \square

Theorem 25. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is a frequency ordering of the Fourier basis. Then there is a constant $C_1 > 0$ such that for all $\epsilon \in (0, 1/2]$ and $N \in \mathbb{N}$*

$$\mu(\pi_N U) \leq \frac{C_1 \epsilon^{1/3}}{N^{2/3}}.$$

Furthermore, there is a constant $C_2 > 0$ such that for all $\epsilon \in (0, 1/2]$ there exists an $M(\epsilon) \in \mathbb{N}$ such that for all $N \geq M$ we have the bound

$$\mu(\pi_N U) \geq \frac{C_2 \epsilon^{1/3}}{N^{2/3}}.$$

Proof. Upper Bound: Without loss of generality we can assume τ is the natural ordering of B_p . Recall that from (51) we have

$$|U_{m,n}|^2 = \frac{n-1/2}{|\lambda \circ \rho(m)|} J_{n-1/2}^2(2\pi\epsilon\lambda \circ \rho(m)) \quad (60)$$

$$= 4\epsilon(n-1/2) j_{n-1}^2(2\pi\epsilon\lambda \circ \rho(m)). \quad (61)$$

We shall first derive two useful bounds; notice that if we apply (54) to (61) then we get the bound, for some constant $\beta > 0$,

$$|U_{m,n}|^2 \leq 4\epsilon(n-1/2) \cdot (\beta(n-1/2)^{-5/6})^2 \leq 4\epsilon\beta^2(n-1/2)^{-2/3}. \quad (62)$$

Secondly we shall use the following inequality from [24]

$$|J_\nu(x)| \leq b\nu^{-1/3} \quad \nu > 0, \quad x \in \mathbb{R}, \quad (63)$$

where $b > 0$ is some constant. Applying this to (60) gives the bound

$$|U_{m,n}|^2 \leq \frac{n-1/2}{|\lambda \circ \rho(m)|} (b(n-1/2)^{-1/3})^2 \leq \frac{b^2(n-1/2)^{1/3}}{|\lambda \circ \rho(m)|}. \quad (64)$$

Recall that our goal is to estimate $|U_{m,n}|$ uniformly in n as $m \rightarrow \infty$. We first apply the case $n - 1/2 \geq \epsilon|\lambda \circ \rho(m)|$ to (62) to give the bound

$$|U_{m,n}|^2 \leq 4\epsilon\beta^2(\epsilon\lambda \circ \rho(m))^{-2/3} \leq \frac{4\beta^2\epsilon^{1/3}}{|\lambda \circ \rho(m)|^{2/3}}.$$

For the other case $n - 1/2 \leq \epsilon|\lambda \circ \rho(m)|$ we use (64) to give the bound

$$|U_{m,n}|^2 \leq \frac{b^2(\epsilon|\lambda \circ \rho(m)|)^{1/3}}{|\lambda \circ \rho(m)|} \leq \frac{b^2\epsilon^{1/3}}{|\lambda \circ \rho(m)|^{2/3}} = \frac{b^2\epsilon^{1/3}2^{2/3}}{(m-1)^{2/3}},$$

which gives a global upper bound in terms of $m \geq 2$ and $\epsilon \in (0, 1/2]$. If $m = 0$, i.e. $\lambda \circ \rho(m) = 0$, then since $j_n(0) = 0$ for $n \geq 1$ (see (48)) we deduce that $\mu(\pi_1 U) = \epsilon|j_0(0)|^2 = \epsilon$ which is a stronger bound than required.

Lower Bound: By (56) we know that

$$j'_{n+1/2,1} - j'_{n-1/2,1} \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (65)$$

With this in mind let $n(m) \in \mathbb{N}$ denote the nearest $j'_{n-1/2,1}$ to $|2\pi\epsilon\lambda \circ \rho(m)|$. From (65) we observe

$$|j'_{n(m)-1/2,1} - |2\pi\epsilon\lambda \circ \rho(m)|| \leq 1/2 + \eta(m, \epsilon), \quad (66)$$

where η is such that $\eta(m, \epsilon) \rightarrow 0$ as $m \rightarrow \infty$ for any fixed ϵ . By using the same method as in (58) we find that

$$\begin{aligned} & |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1}) - J_{n(m)-1/2,1}(|2\pi\lambda \circ \rho(m)|)| \\ & \leq \frac{|j'_{n(m)-1/2,1} - |2\pi\epsilon\lambda \circ \rho(m)||^2}{2} \cdot |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})| \\ & \leq 2^{-1} \cdot (2^{-1} + \eta(m, \epsilon))^2 \cdot |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})| \\ & = \xi(m, \epsilon) \cdot |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})|. \end{aligned}$$

Where $\xi(m, \epsilon) \rightarrow 8^{-1}$ as $m \rightarrow \infty$ with ϵ fixed. This tells us that

$$\begin{aligned} & |J_{n(m)-1/2,1}(2\pi\lambda \circ \rho(m))| = |J_{n(m)-1/2,1}(2\pi\epsilon\lambda \circ \rho(m))| \\ & \geq |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})| - |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1}) - J_{n(m)-1/2,1}(|2\pi\lambda \circ \rho(m)|)| \\ & \geq (1 - \xi(m, \epsilon))|J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})|. \end{aligned}$$

Combining this with (60) we see that, using (57),

$$\begin{aligned} |U_{m,n(m)}|^2 &= \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} |J_{n(m)-1/2}^2(2\pi\epsilon\lambda \circ \rho(m))| \\ &\geq \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} \cdot (1 - \xi(m, \epsilon))^2 \cdot |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})|^2 \\ &\geq \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} \cdot (1 - \xi(m, \epsilon))^2 \cdot (\kappa(n(m) - 1/2)^{-1/3} + \mathcal{O}((n(m) - 1/2)^{-1}))^2. \end{aligned} \quad (67)$$

By (56), (66) and the fact that ρ is a standard ordering we know that (for ϵ fixed)

$$\frac{n(m)}{|\pi\epsilon m|} \rightarrow 1, \quad \text{as } m \rightarrow \infty.$$

Therefore we know that there is an $M(\epsilon) \in \mathbb{N}$ and a constant $C > 0$ such that for all $m \geq M$ and $\epsilon \in (0, 1/2]$ we have

$$|U_{m,n(m)}|^2 \geq C \cdot \epsilon^{1/3} \cdot m^{-2/3}.$$

Consequently for $N \geq M(\epsilon)$ we have $\mu(\pi_N U) \geq C \cdot \epsilon^{1/3} \cdot N^{-2/3}$. \square

Summarising our results in this section, while throwing away ϵ dependence again, we deduce the following

Corollary 26. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is a frequency ordering, τ is the natural ordering and $\epsilon \in (0, 0.45]$. Then*

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-2/3}). \quad (68)$$

8 Asymptotic Incoherence and Subsampling Strategies

We have shown that there is faster asymptotic incoherence for the Fourier-wavelet case than for the Fourier-polynomial case. We shall demonstrate how this difference is vital for choosing an effective sampling strategy.

Consider the problem of reconstructing the function $f \in L^2[-1, 1]$ from its samples $\{\langle f, g \rangle : g \in B_f(1/2)\}$, where f is defined as

$$f(x) = (1 - \cos(8\pi x)) \cdot \mathbb{1}_{[0,1]}(x), \quad x \in [-1, 1]. \quad (69)$$

The function f is reconstructed as follows: Let $U := [(B_f(2^{-1}), \rho), (B_2, \tau)]$ for some orderings ρ, τ and a reconstruction basis B_2 . The number 2^{-1} is present here to ensure the span of B_f contains $L^1[-1, 1]$. It is assumed that ρ is a frequency ordering. Next let $\Omega \subset \mathbb{N}$ denote the set of subsamples from $B_f(2^{-1})$ (indexed by ρ), P_Ω the projection operator onto Ω and $\hat{f} := (\langle f, \rho(m) \rangle)_{m \in \mathbb{N}}$. We then attempt to approximate f by $\sum_{n=1}^{\infty} \tilde{x}_n \tau(n)$ where $\tilde{x} \in \ell^1(\mathbb{N})$ solves the optimisation problem

$$\min_{x \in \ell^1(\mathbb{N})} \|x\|_1 \quad \text{subject to} \quad P_\Omega U x = P_\Omega \hat{f}. \quad (70)$$

Since the optimisation problem is infinite dimensional we cannot solve it numerically so instead we proceed as in [3] and truncate the problem, approximating f by $\sum_{n=1}^R \tilde{x}_n \tau(n)$ (for $R \in \mathbb{N}$ large) where $\tilde{x} = (\tilde{x}_n)_{n=1}^R$ now solves the optimisation problem

$$\min_{x \in \mathbb{C}^R} \|x\|_1 \quad \text{subject to} \quad P_\Omega U P_R x = P_\Omega \hat{f}. \quad (71)$$

We shall be using the SPGL1 package [6] to solve (71) numerically. We focus on two choices of reconstruction bases:

1. $B_2 = B_{\text{bw}}$ with Daubechies4 boundary wavelets, τ is a leveled ordering.
2. $B_2 = B_p$ with Legendre polynomials, τ is a natural ordering.

The coefficients of the decomposition of f into these two bases is shown in Figure 3. The coefficients in the polynomial expansion decay quickly, but there is little sparsity in the first 40 coefficients. On the other hand in the wavelet expansion there is large number of zeros in the first block of coefficients. This, combined with asymptotic incoherence, will enable us to subsample.

We shall be looking at two simple subsampling patterns and how they perform for each reconstruction basis. We shall be subsampling from the first 501 coefficients, and since ρ is a frequency ordering this means that these coefficients correspond to

$$\{\lambda \circ \rho(m) : m = 1, \dots, 501\} = \{-250, -249, \dots, 249, 250\}.$$

If we were to sample all the 501 coefficients then we would achieve a highly accurate reconstruction from both bases⁷. We now consider two subsampling patterns, denoted as pattern A and pattern B which are presented in Figure 4, and now try to use them to reconstruct in the bases B_{bw}, B_p . Pattern A takes all its samples from the first 101 coefficients and there is very little subsampling in this range.

Figure 3: Coefficients of f when decomposed into different reconstruction bases.

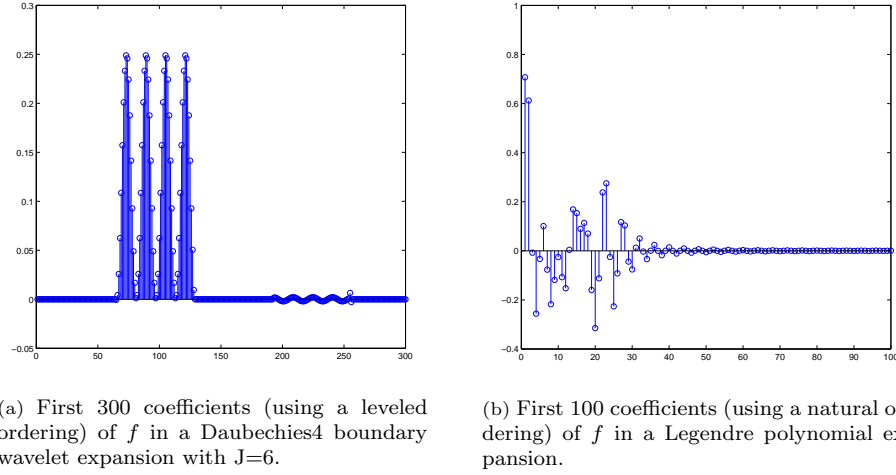
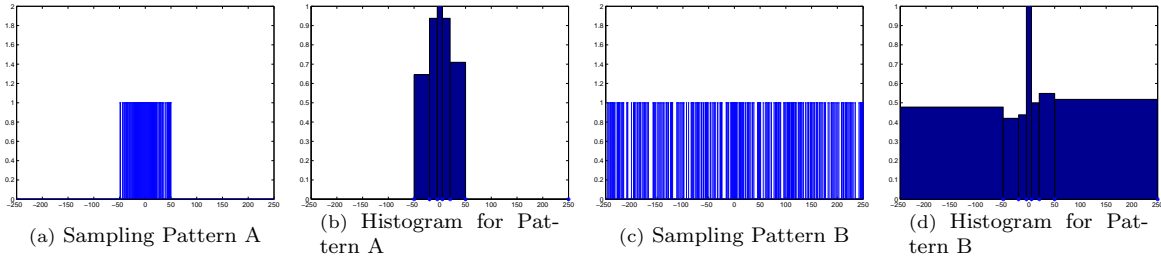


Figure 4: Two sampling patterns and their corresponding histograms.



On the other hand pattern B takes around 50% of the samples from across the first 501 coefficients. Both patterns are constructed by uniformly subsampling in levels.

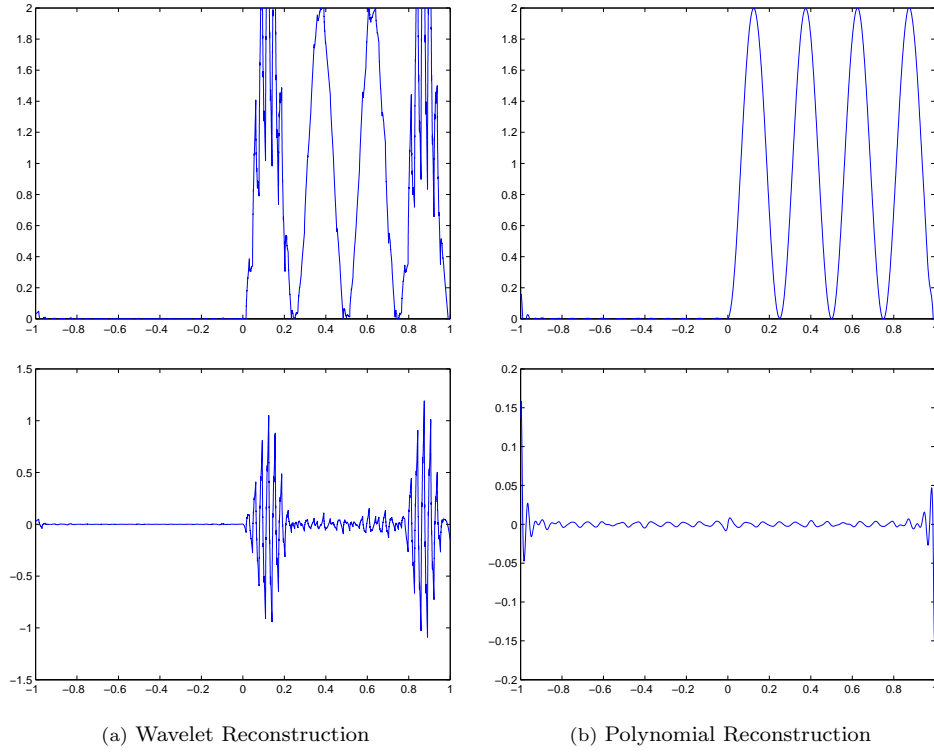
Let us first consider what happens when we use subsampling pattern A, which is shown in Figure 5. We first look at the wavelet reconstruction, which has an L^1 error of 1.52×10^{-1} . The reconstruction fails to reconstruct the smoothness of f , with the first and fourth peaks being particularly jagged. Next consider the polynomial reconstruction, which has an L^1 error of 8.68×10^{-3} . Since polynomials provide a relatively good linear approximation to f , it is unsurprising that using a near full-sampling subsampling pattern for the first 101 Fourier coefficients would give a reasonable reconstruction.

Next we turn to reconstructing f using sampling pattern B. Reconstructions are given in Figure 6. First we look at the wavelet reconstruction which has an L^1 error of 7.14×10^{-3} . Since the wavelet basis expansion of f is sparse and we have asymptotic incoherence, we see that we can obtain a good wavelet reconstruction by subsampling roughly 50% of the 501 Fourier samples. Finally we consider the polynomial reconstruction, with an L^1 error of 7.29×10^{-1} . Due to poor sparsity and slow asymptotic incoherence, subsampling fails to be successful.

This therefore demonstrates that a subsampling pattern should not only be dependent on the function that we are trying to reconstruct, but also on the reconstruction bases that we are using. We must

⁷For all our reconstructions we will be using $R = 1024$.

Figure 5: Reconstructions from Pattern A (above) with errors (below).



stress here that the ability to find two subsampling patterns, where each gives a better reconstruction in a different basis, relies crucially on the different incoherence structures of the two reconstruction problems and not simply the sparsity structure when decomposed into the two reconstruction bases; the same phenomenon can also be demonstrated if we remove f completely and instead fix the sparsity structure (which means solving for a fixed \tilde{x} in our optimisation setup). Asymptotic incoherence not only facilitates subsampling but also allows us to investigate the link between good subsampling patterns and reconstruction bases.

9 Optimality

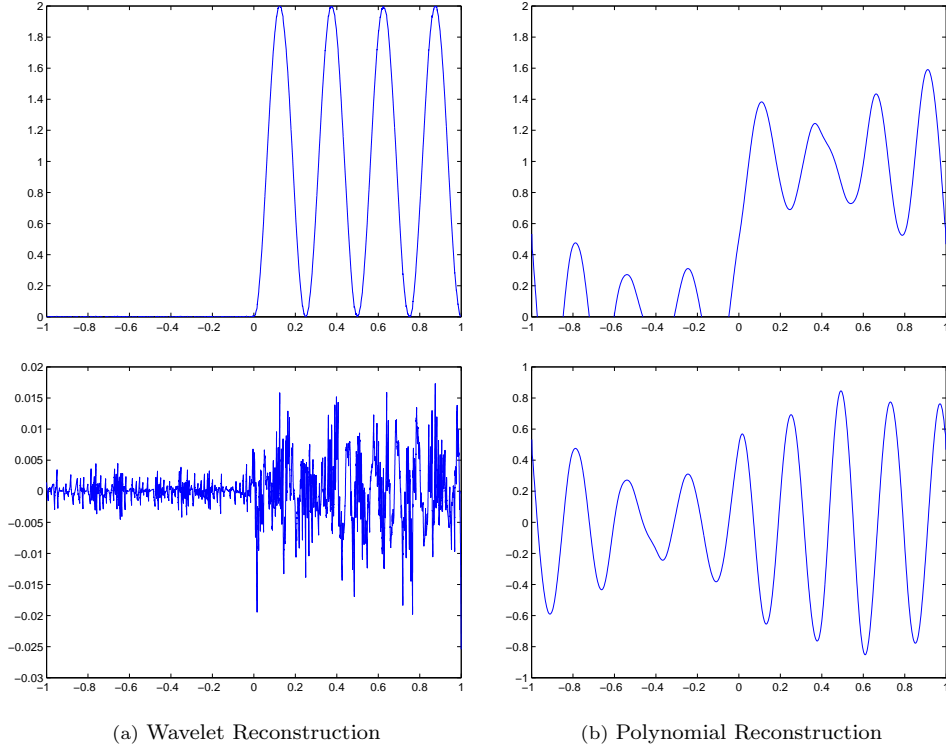
In Sections 6 and 7 we derived bounds on the line coherences for the Fourier-wavelet case (say $U = U_w$) the Fourier polynomial case (say $U = U_p$)

$$\begin{aligned} \mu(\pi_N U_w), \mu(U_w \pi_N) &= \Theta(N^{-1}), & \mu(\pi_N U_p), \mu(U_p \pi_N) &= \Theta(N^{-2/3}), \\ \Rightarrow \mu(R_N U_w), \mu(U_w R_N) &= \Theta(N^{-1}), & \mu(R_N U_p), \mu(U_p R_N) &= \Theta(N^{-2/3}). \end{aligned} \quad (72)$$

This required that we work with a frequency ordering for the Fourier basis and leveled/natural orderings for the wavelet/polynomial basis. Our next goal is to show that no other orderings can improve upon the decay rates in (72).

Since we want to compare decay rates with different orderings, we need a precise way of saying one ordering has a slower decay rate than another:

Figure 6: Reconstructions from Pattern B with errors.



Definition 27 (Relations on the set of orderings). Let $U_1 := [(B_1, \rho_1), (B_2, \tau)]$, $U_2 := [(B_1, \rho_2), (B_2, \tau)]$. If

$$\mu(R_N U_1) = \mathcal{O}(\mu(R_N U_2)), \quad N \rightarrow \infty, \quad (73)$$

then we write $\rho_1 \prec \rho_2$ and say that ' ρ_1 has a faster decay rate than ρ_2 for the basis pair (B_1, B_2) '. If also $\rho_2 \prec \rho_1$ we write $\rho_1 \sim \rho_2$. These relations, defined on the set of orderings of B_1 which we shall denote as $\mathcal{R}(B_1)$, depend only on the basis pair (B_1, B_2) , and are therefore independent of τ .

Notice that \prec is a reflexive transitive relation on $\mathcal{R}(B_1)$ and \sim is an equivalence relation on $\mathcal{R}(B_1)$. Furthermore, we can use the relation to define a partial order on the equivalence classes of $\mathcal{R}(B_1)$ by the definition

$$[a] \prec [b] \iff a \prec b,$$

where $[a]$ denotes the equivalence class containing a . Furthermore, we say an equivalence class $[a]$ is 'optimal' if we have

$$[a] \prec [b], \quad \forall b \in \mathcal{R}(B_1).$$

Definition 28 (Optimal ordering). Given the setup above, then any element of the optimal equivalence class is called an 'optimal ordering of the basis pair (B_1, B_2) '.

It shall be shown in Lemma 38 that optimal orderings always exist. Notice that ρ is an optimal ordering if and only if for every other ordering ρ' we have $\rho \prec \rho'$. An optimal ordering has a corresponding optimal decay rate.

Definition 29 (Optimal decay rate). Suppose $\rho : \mathbb{N} \rightarrow B_1$ is an optimal ordering for the basis pair (B_1, B_2) and $U = [(B_1, \rho), (B_2, \tau)]$. Then any decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ which satisfies $f(N) = \Theta(\mu(R_N U))$ is said to represent the 'optimal decay rate' of the basis pair (B_1, B_2) .

At this point it is worth explaining why we define \prec in terms of block coherences and not line coherences. This is because $\mu(\pi_N U_1) = \mathcal{O}(\mu(\pi_N U_2))$ implies $\mu(R_N U_1) = \mathcal{O}(\mu(R_N U_2))$ but not the other way around. Furthermore, $\mu(\pi_N U_1)$ is often not a decreasing function of N and a statement such as $\mu(\pi_N U_1) = \Theta(f(N))$ for a decreasing function f is not possible. However, the case when this does hold is very important to us:

Definition 30 (Optimal In lines). *Let $U = [(B_1, \rho), (B_2, \tau)]$. If $\mu(\pi_N U) = \Theta(f(N))$ for a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ then ρ is said to be ‘optimal in lines’ for the basis pair (B_1, B_2) .*

From (72) we see that the orderings used in these cases (frequency/leveled/natural) are optimal in lines by definition. We now show that the use of the word ‘optimal’ here is justified:

Proposition 31. *Let $U_1 = [(B_1, \rho_1), (B_2, \tau)]$ and $U_2 = [(B_1, \rho_2), (B_2, \tau)]$. If there exists a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ such that*

$$f(N) \leq \mu(\pi_N U_1), \quad N \in \mathbb{N}, \quad (74)$$

then $f(N) \leq \mu(R_N U_2)$ for every $N \in \mathbb{N}$.

Proof. Let $\theta(N)$ denote the smallest $m \in \mathbb{N}$ such that $\rho_1(m) \in \{\rho_2(k)\}_{k=N}^\infty$ and $m' = m'(N) \in \{N, N+1, \dots\}$ be such that $\rho_1(\theta(N)) = \rho_2(m'(N))$. Now notice that $\theta(N) \leq N$ since $\{\rho_2(k)\}_{k=N}^\infty$ can only miss at most the first $N-1$ of the $\rho_1(k)$ ’s. Combining this with the fact that f is decreasing we see that

$$f(N) \leq f(\theta(N)) \leq \mu(\pi_{\theta(N)} U_1) = \mu(\pi_{m'(N)} U_2) \leq \mu(R_N U_2).$$

□

Corollary 32. *Let $U_1 = [(B_1, \rho_1), (B_2, \tau)]$ and $U_2 = [(B_1, \rho_2), (B_2, \tau)]$. If ρ_1 is optimal in lines, i.e. $\mu(\pi_N U_1) = \Theta(f(N))$ for a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$, then $f(N) = \mathcal{O}(\mu(R_N U_2))$ and $\rho_1 \prec \rho_2$. Consequently, ρ_1 is optimal.*

Proof. By definition $\mu(\pi_N U_1) = \Theta(f(N))$ implies $C \cdot f(N) \leq \mu(\pi_N U_1)$ for some constant $C > 0$. Applying the proposition gives us $C \cdot f(N) \leq \mu(R_N U_2)$, i.e. $f(N) = \mathcal{O}(\mu(R_N U_2))$.

Recall from Lemma 11 that $\mu(\pi_N U_1) = \Theta(f(N))$ implies $\mu(R_N U_1) = \Theta(f(N))$. Therefore $\mu(R_N U_1) = \mathcal{O}(\mu(R_N U_2))$ which by definition means $\rho_1 \prec \rho_2$.

□

We can now use Corollary 32 to immediately deduce the bounds in (72) are optimal for their respective basis pairs:

Theorem 33. *1. **Fourier-Wavelet Case:** Let $\epsilon \in I_{J,p}$. Frequency orderings are optimal for the basis pair $(B_f(\epsilon), B_w)$. Leveled orderings are optimal for the basis pair $(B_w, B_f(\epsilon))$. In both cases the optimal decay rate is $\Theta(N^{-1})$. These statements still hold with B_w replaced with the boundary wavelet basis B_{bw} and $\epsilon \in I_{J,p}$ by $\epsilon \in (0, 1/2]$.*

*2. **Fourier-Polynomial Case:** Let $\epsilon \in (0, 1/2]$. Frequency orderings are optimal for the basis pair $(B_f(\epsilon), B_p)$. Leveled orderings are optimal for the basis pair $(B_p, B_f(\epsilon))$. In both cases the optimal decay rate is $\Theta(N^{-2/3})$.*

10 Theoretical Limits

We now look at the general abstract case where U is an isometry and ask; is there a universal lower bound on the block coherences?

Theorem 34. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then $\sum_N \mu(R_N U)$ diverges.*

Proof. Suppose that $\sum_N \mu(R_N U)$ converges, Then, we can find $N' \in \mathbb{N}$ such that $\sum_{N=N'}^\infty \mu(R_N U) \leq 1/4^2$. Therefore if we write $U = (u_{i,j})_{i,j \in \mathbb{N}}$ then

$$\sum_{N=N'}^\infty |u_{N,j}|^2 \leq \sum_{N=N'}^\infty \mu(R_N U) \leq 1/4^2, \quad j \in \mathbb{N}. \quad (75)$$

Now define the vectors

$$v_j := (u_{i,j})_{i \in \mathbb{N}}, \quad v_j^1 := (u_{i,j})_{i=1}^{N'-1}, \quad v_j^2 := (u_{i,j})_{i=N'}^\infty, \quad j \in \mathbb{N}.$$

Inequality (75) says that $\|v_j^2\|_2 \leq 1/4$ for every $j \in \mathbb{N}$. Since U is an isometry, we know its columns are normalised, i.e. $\|v_j\|_2 = 1$, and so we deduce $\|v_j^1\|_2 \geq 3/4$ for every $j \in \mathbb{N}$. Let $w_j := v_j^1 / \|v_j^1\|_2$, $j \in \mathbb{N}$. Since the $w_j \in \mathbb{C}^{N'-1}$ are all finite dimensional we claim that

$$\sup_{\substack{j,j' \in \{1, \dots, M\} \\ j \neq j'}} |\langle w_j, w_{j'} \rangle| \rightarrow 1, \quad \text{as } M \rightarrow \infty. \quad (76)$$

To see this, notice that for every $\epsilon > 0$, there exists a $\delta > 0$, such that for all $j \in \mathbb{N}$ the set $W_j(\epsilon) := \{w \in \mathbb{C}^{N'-1} : |\langle w_j, w \rangle| > 1 - \epsilon\}$ contains the open set $B_\delta(w_j)$ of radius δ centered at w_j . It must be the case that there are $j_1, j_2 \in \mathbb{N}$, $j_1 \neq j_2$ such that $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \emptyset$, else the union

$$\bigcup_{j \in \mathbb{N}} B_{\delta/2}(w_j) \cup \bigcup_{\substack{w \notin \bigcup_{j \in \mathbb{N}} B_{\delta/2}(w_j) \\ w \in \mathbb{C}^{N'-1}}} B_{\delta/4}(w),$$

would form an open cover of the unit ball in $\mathbb{C}^{N'-1}$ with no finite subcover⁸, contradicting compactness of the unit ball in $\mathbb{C}^{N'-1}$. Since $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \emptyset$, $w_{j_1} \in B_\delta(w_{j_2}) \subset W_{j_2}(\epsilon)$ and so $|\langle w_{j_1}, w_{j_2} \rangle| > 1 - \epsilon$. Since $\epsilon > 0$ was arbitrary we have proved (76).

Therefore, by (76) we know there exists $j_1, j_2 \in \mathbb{N}$, $j_1 \neq j_2$ such that $|\langle w_{j_1}, w_{j_2} \rangle| > 1/2$ and therefore we deduce that

$$|\langle v_{j_1}^1, v_{j_2}^1 \rangle| > \frac{1}{2} \|v_{j_1}^1\|_2 \|v_{j_2}^1\|_2 > \frac{3^2}{2 \cdot 4^2}. \quad (77)$$

Furthermore, since $\|v_{j_1}^2\|_2, \|v_{j_2}^2\|_2 \leq 1/4$ we know that

$$|\langle v_{j_1}^2, v_{j_2}^2 \rangle| \leq \|v_{j_1}^2\|_2 \|v_{j_2}^2\|_2 \leq \frac{1}{4^2}. \quad (78)$$

Therefore, combining (77) with (78) gives us

$$\begin{aligned} |\langle v_{j_1}, v_{j_2} \rangle| &= |\langle v_{j_1}^1, v_{j_2}^1 \rangle + \langle v_{j_1}^2, v_{j_2}^2 \rangle| \geq |\langle v_{j_1}^1, v_{j_2}^1 \rangle| - |\langle v_{j_1}^2, v_{j_2}^2 \rangle| \\ &\geq \frac{3^2}{2 \cdot 4^2} - \frac{1}{4^2} = \frac{7}{2 \cdot 4^2} > 0. \end{aligned}$$

However, since U is an isometry and $j_1 \neq j_2$, we know that $\langle v_{j_1}, v_{j_2} \rangle = 0$ and therefore we have a contradiction. \square

Corollary 35. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then there does not exist an $\epsilon > 0$ such that*

$$\mu(R_N U) = \mathcal{O}(N^{-1-\epsilon}), \quad N \rightarrow \infty.$$

Noting the above corollary and that $\mu(W) \geq N^{-1}$ is the best lower bound possible for any finite isometry $W \in \mathbb{C}^N \times \mathbb{C}^N$ it might be tempting to believe $\mu(R_N U) = \Theta(N^{-1})$ is the best decay rate we can achieve for an isometry $U \in \mathcal{B}(l^2(\mathbb{N}))$. However, it turns out that Theorem 34 cannot be improved without imposing additional conditions on U :

⁸Any finite subcover would miss infinitely many of the points w_j .

Theorem 36. *Let $f, g : \mathbb{N} \rightarrow \mathbb{R}$ be any two strictly positive decreasing functions and suppose that $\sum_N f(N)$ diverges. Then there exists $U \in \mathcal{B}(l^2(\mathbb{N}))$ an isometry with*

$$\mu(R_N U) \leq f(N), \quad \mu(U R_N) \leq g(N), \quad N \in \mathbb{N}. \quad (79)$$

Proof. The proof is constructive. We may assume without loss of generality that $f(N), g(N) \leq 1$ for all $N \in \mathbb{N}$. We will construct a matrix $U = (u_{i,j})_{i,j \in \mathbb{N}}$ satisfying (79) with normalised columns, $v_j := (u_{i,j})_{i \in \mathbb{N}}, j \in \mathbb{N}$, having disjoint support. With this in mind we partition \mathbb{N} as follows:

$$\mathbb{N} = \bigcup_{i=1}^{\infty} \Omega_i, \quad \Omega_i := 2^{i-1}\mathbb{N} \setminus 2^i\mathbb{N}.$$

Let $j \in \mathbb{N}$ be fixed and define recursively (for⁹ $N \in \mathbb{N}$)

$$(v_j)_N = \begin{cases} (g(j)f(N))^{1/2}, & \text{if } \sum_{i=1}^{N-1} ((v_j)_i)^2 + g(j)f(N) \leq 1, \quad N \in \Omega_j, \\ (1 - \sum_{i=1}^{N-1} ((v_j)_i)^2)^{1/2}, & \text{if } \sum_{i=1}^{N-1} ((v_j)_i)^2 \leq 1, \\ \sum_{i=1}^{N-1} ((v_j)_i)^2 + g(j)f(N) \geq 1, \quad N \in \Omega_j, \\ 0, & \text{Otherwise.} \end{cases} \quad (80)$$

It is immediate from the definition that v_j is supported on Ω_j and $((v_j)_N)^2 \leq f(N)g(j)$ for every $N, j \in \mathbb{N}$ which implies that (79) holds. Furthermore, it is easy to show by induction on N that $\|v_j\|_2 \leq 1$. Since f is decreasing and by the structure of the set Ω_j , $\sum_{N \in \Omega_j} f(N)$ diverges for every j and consequently there is an $N' \in \mathbb{N}$ such that

$$\sum_{\substack{N \in \Omega_j \\ N \leq N'}} g(j)f(N) \geq 1, \quad \sum_{\substack{N \in \Omega_j \\ N \leq N'-1}} g(j)f(N) \leq 1.$$

For $N \leq N' - 1, N \in \Omega_j$ we fall into the first case of (80), however for $N = N'$ we fall into case 2, and therefore $\sum_{i=1}^{N'} ((v_j)_i)^2 = 1$. This means $\|v_j\|_2 = 1$ for every j and consequently U is an isometry. \square

Although this negative result shows that we cannot define an analogue of perfect incoherence for asymptotic incoherence, if we restrict our decay function to be a power law, i.e. $f(N) := CN^{-\alpha}$ for some constants $\alpha, C > 0$ then the largest possible value of $\alpha > 0$ such that (79) holds for an isometry U is $\alpha = 1$, which is what we achieved in the Fourier-wavelet case.

11 Alternative Notions of Optimality

Before we finish, we would like to discuss further why we work with Definition 27 as our notion of optimal decay and discuss a possible alternative. One argument against the definition of optimality we use is that it is only unique up to constants, since it relies only on order notation. This is somewhat inconvenient if one wants to work with concrete estimates. Therefore it may be tempting to strengthen the notion of optimality in some way. One possible alternative would be the following:

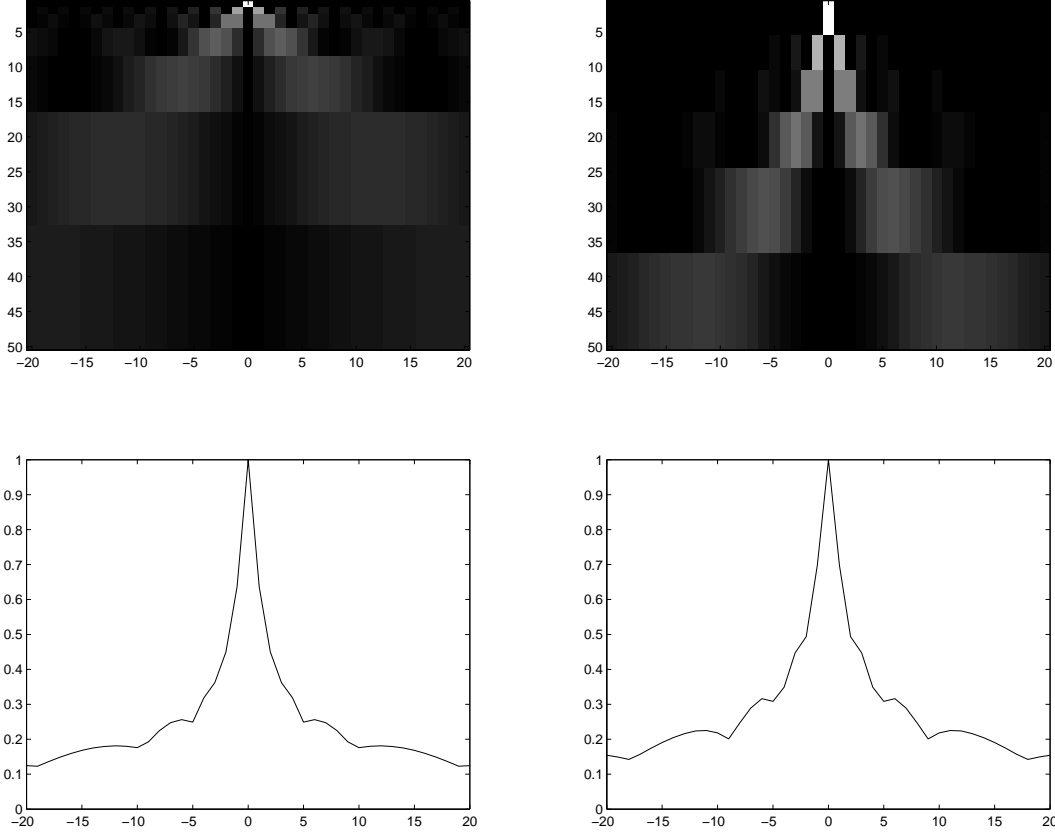
Definition 37 (Best ordering). *Let (B_1, B_2) be a basis pair. Then any ordering $\rho : \mathbb{N} \rightarrow B_1$ is said to be a ‘best ordering’ if for any other ordering τ of B_2 and $U = [(B_1, \rho), (B_2, \tau)]$ we have that the function $g(N) := \mu(\pi_N U)$ is decreasing.*

Notice that for a best ordering we have $\mu(\pi_N U) = \mu(R_N U)$. If ρ' is any other ordering and $U' = [(B_1, \rho'), (B_2, \tau)]$ then since $R_N U'$ must contain one of the first N lines of U we must have that

$$\mu(R_N U') \geq \min_{M=1, \dots, N} \mu(\pi_M U) \geq \mu(\pi_N U) = \mu(R_N U),$$

and we deduce that $\rho < \rho'$. This shows that any best ordering is optimal.

⁹Here we use the convention that $\sum_{i=1}^{N-1}$ is an empty sum if $N = 1$.



(a) Incoherence matrix and column maxima for a Haar wavelet basis (with Fourier).

(b) Incoherence matrix and column maxima for Daubechies6 wavelet basis.

Figure 7: Here are two (20×20) centrally truncated wavelet-Fourier Incoherence matrices (brighter means larger absolute value) and their corresponding column maxima. The columns denote the Fourier basis (viewed as \mathbb{Z}) and the rows denote the wavelet basis (ordered top to bottom). Notice that there is a slight difference in the best orderings (by looking around $-10, +10$ on the horizontal axis) even though the general decay rate is similar. The maxima are taken over a much larger matrix to ensure accuracy.

Lemma 38. *Suppose that we have a basis pair (B_1, B_2) . Then one of the following two results must hold:*

- (1) *There is at least one best ordering.*
- (2) *Every ordering of B_1 is optimal for (B_1, B_2) .*

Proof. Let $\rho : \mathbb{N} \rightarrow B_1$, $\tau : \mathbb{N} \rightarrow B_2$ be any orderings of B_1, B_2 respectively and $U = [(B_1, \rho), (B_2, \tau)]$. Now first assume that for any finite subset $D \subset \mathbb{N}$

$$\sup_{N \in \mathbb{N} \setminus D} \mu(\pi_N U), \quad (81)$$

is attained for some $N \in \mathbb{N} \setminus D$. In this case we can then construct a best ordering $\rho^* : \mathbb{N} \rightarrow B_1$ inductively by letting (for $N = 1$)

$$\rho^*(1) \in \operatorname{argmax}_{f \in B_1} \sup_{n \in \mathbb{N}} |\langle \tau(n), f \rangle|,$$

and for $N \geq 2$ we set

$$\rho^*(N) \in \underset{f \in B_1}{\operatorname{argmax}} \sup_{n \in \mathbb{N}} |\langle \tau(n), f \rangle|.$$

$f \notin \{\rho^*(1), \dots, \rho^*(N-1)\}$

Note that it is clear from the construction that this is an actual ordering. Therefore if our original assumption holds we conclude that 1) must hold too. If our assumption does not hold this means there exists a finite subset $D \subset \mathbb{N}$ such that the supremum (81) is not attained for any $N \in \mathbb{N} \setminus D$. This means that if we remove finitely many elements from $\mathbb{N} \setminus D$ the supremum will remain unchanged. Therefore if N' is the largest natural number in D we find that

$$\mu(R_M U) = \sup_{N \in \mathbb{N} \setminus D} \mu(\pi_N U), \quad \forall M > N',$$

and so $\mu(R_N U)$ is eventually constant as a function of N . This means that for *any* ordering ρ' of B_1 and $U' = [(B_1, \rho'), (B_2, \tau)]$, $\mu(R_N U')$ is eventually constant. It follows that any two orderings of B_1 are equivalent under \sim and consequently 2) holds. \square

Lemma 39. *Suppose that we have a basis pair (B_1, B_2) with two orderings $\rho : \mathbb{N} \rightarrow B_1$, $\tau : \mathbb{N} \rightarrow B_2$ of B_1, B_2 respectively. If $U = [(B_1, \rho), (B_2, \tau)]$ satisfies*

$$\mu(\pi_N U) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

then a best ordering exists.

Proof. The supremum (81) is always attained and therefore we fall into case 1) of the previous lemma. \square

These two results tell us that optimal orderings always exist and best orderings exist in cases where we expect decay in the line/block coherences, i.e. every case where we want to study coherence decay.

Therefore, why did we not work with the definition of best ordering as our notion of optimality instead? The answer to this question is that best orderings are more exotic and far less simple to describe than optimal orderings. Figure 7 shows the Fourier-wavelet case where the best orderings are wavelet-dependent, even though we can describe optimal orderings in a wavelet-independent manner using frequency orderings. Although the difference between best orderings is very minor in Figure 7, this difference becomes considerable when working with the higher dimensional Fourier-wavelet cases.

12 Outlook and Future Work

The results presented here can be extended to higher dimensional cases, but the complexity increases significantly and is the subject of future work. Unlike the one-dimensional Fourier/wavelet case we covered here, the optimal orderings in the multidimensional Fourier/(separable) wavelet case can be wavelet dependent in more than three dimensions however the optimal decay rates are the same ($\Theta(N^{-1})$). These results rely heavily on the optimality theory developed here. An alternative direction to develop upon this work would be to cover other reconstruction bases other than the wavelet and polynomial bases presented here.

13 Acknowledgements

A. D. Jones acknowledges EPSRC grant EP/H023348/1. B. Adcock acknowledges NSF DMS grant 1318894. A. C. Hansen acknowledges support from a Royal Society University Research Fellowship as well as EPSRC grant EP/L003457/1.

References

- [1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, Washington, D.C., 1964.
- [2] B. Adcock, A. Bastounis, A. C. Hansen, and B. Roman. On fundamentals of models and sampling in compressed sensing. *Preprint*, 2015.
- [3] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, 2015.
- [4] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Preprint*, 2015.
- [5] B. Adcock, A. C. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *Adv. in Imag. and Electr. Phys.*, 182:187279, 2014.
- [6] E. Berg and M. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal of Scientific Computing*, 31(2):890–912, 2008.
- [7] C. Boyer, P. Weiss, and J. Bigot. An algorithm for variable density sampling with block-constrained acquisition. *Siam Journal of Imaging Sciences*, 7:1080–1107, 2014.
- [8] E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.
- [9] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.
- [10] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [11] 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.
- [12] A. Cohen, I. Daubechies, and P. Vial. Wavelet bases on the interval and fast algorithms. *Journal of Applied and Computational Harmonic Analysis*, 1993.
- [13] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41(7):909–996, 1988.
- [14] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.
- [15] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [16] Y. C. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.
- [17] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.
- [18] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhuser, 2013.
- [19] T. N. G. Strang. Wavelets & filter banks. *Wellesley-Cambridge Press*, 1996.

- [20] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363386, 2015.
- [21] E. Hernández and G. Weiss. *A first course on wavelets*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1996. With a foreword by Yves Meyer.
- [22] F. Krahmer, H. Rauhut, and R. Ward. Local coherence sampling in compressed sensing. *Proceedings of the 10th International Conference on Sampling Theory and Applications*, 2013.
- [23] F. Krahmer and R. Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Process.*, 23 (2):612–622, 2014.
- [24] L. J. Landau. Bessel functions: monotonicity and bounds. *J. London Math. Soc. (2)*, 61(1):197–215, 2000.
- [25] 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.
- [26] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131(0):70–91, 2013.
- [27] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: the application of compressed sensing for rapid MRI imaging. *Magn. Reson. Imaging*, 58(6):1182–1195, 2007.
- [28] S. Mallat. *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition, 2009.
- [29] 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.
- [30] H. Rauhut. and R. Ward. Sparse legendre expansions via ℓ_1 -minimization. *Journal of Approximation Theory*, 164:517–533, 2012.
- [31] B. Roman, B. Adcock, and A. Hansen. On asymptotic structure in compressed sensing. *arXiv:1406.4178*, 2014.
- [32] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, Cambridge, England, 1944.