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

A. Jones
University of Cambridge

B. Adcock
Purdue Univesity

A. Hansen
University of Cambridge

Abstract

Recently, it has been shown that incoherence is an unrealistic assumption for compressed sensing when applied to infinite-dimensional inverse problems. Instead, the key property that permits efficient recovery in such problems is so-called asymptotic incoherence. The purpose of this paper is to study this new concept, and its implications towards the design of optimal sampling strategies. We determine how fast the 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 [8] and Donoho [13], has been one of the major achievements in applied mathematics in the last decade [6, 12, 14–16]. 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) [17, 23], X-ray Computed Tomography [9,25], Electron Microscopy [21,22], etc – incoherence is completely lacking. The reason for this can be traced to the observation that classical inverse problems are typically based on the continuous integral transforms of Fourier or Radon type. Via the Fourier slice theorem, the latter can be viewed as a problem of sampling the continuous Fourier transform along radial lines. Hence in both settings, the resulting recovery problem is that of reconstructing an unknown function f from pointwise samples of its Fourier transform. As an inverse problem, we can write this as follows:

$$g = \mathcal{F}f, \quad f \in \mathrm{L}^2(\mathbb{R}^d), \tag{1.1}$$

where we are only given access to a finite set of pointwise values of g. Here

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

denotes the d-dimensional Fourier transform.

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 U. Wavelets, or their various generalizations, are frequently used as the sparsifying transformation, and for smooth functions, one often considers orthogonal polynomials. However, the combination of Fourier samples and wavelet sparsity is completely coherent (see Definition 1.1 for the definition of incoherence). Fortunately, as Figure 1 reveals, although such a measurement matrix is coherent, it is also *asymptotically incoherent*: that is to say, the high coherence (large matrix entries) are isolated to a leading submatrix of U (see Definition 1.4 for a formal definition). This phenomenon has been well documented in [3, 4, 19]. It is precisely this property that allows for the efficient use of compressed sensing in this setting. However, to do this successfully, one must employ sampling strategies that differ substantially from uniform random subsampling, and take into account the local variations in coherence. In other words, to properly understand how to subsample in this setting, it is crucial to estimate the asymptotic coherence. Such estimates are the main topic of this paper.

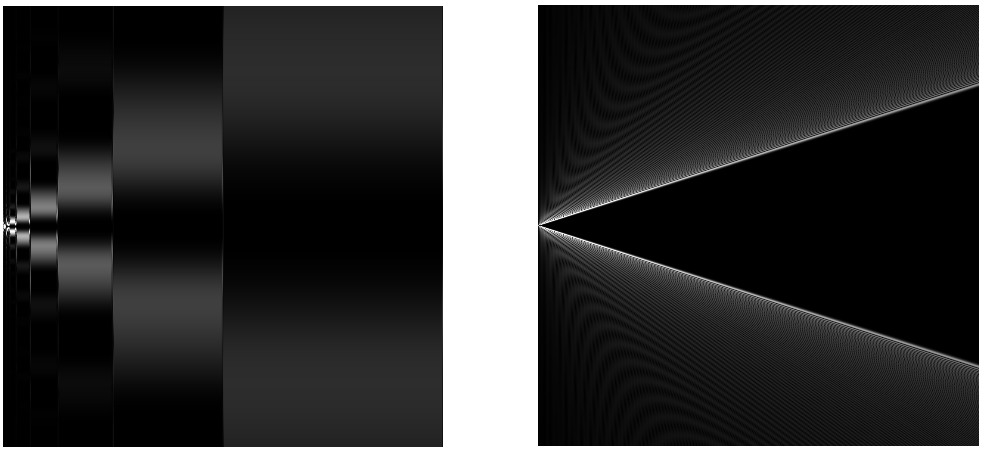


Figure 1: Plots of the absolute values of the entries of the matrix U corresponding to Fourier sampling with Daubechies6 boundary wavelets (left) and Legendre polynomials (right). Light regions correspond to large values and dark regions to small values.

1.1 Compressed Sensing and the Coherence Barrier

Let us now provide some background regarding compressed sensing and incoherence. We commence with the definition of the latter:

Definition 1.1 (Incoherence)**.** *Let $U = (U_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}$ be an isometry. The coherence of U is*

$$\mu(U) = \max_{i,j=1,\dots,N} |U_{ij}|^2 \in [N^{-1}, 1].$$

We say that U is 'perfectly incoherent' if $\mu(U) = N^{-1}$.

Note that the definition of μ obviously extends to the infinite-dimensional case, where U is an isometry of $l^2(\mathbb{N})$. Standard compressed sensing theory 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

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

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

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot \log(\epsilon^{-1}) \cdot \log(N), \tag{1.2}$$

(see [7] and [2])[1].

The estimate (1.2) demonstrates how the three pillars of compressed sensing – sparsity, incoherence and uniform random subsampling – combine to allow for recovery with substantial subsampling. However, now suppose that $\mu(U)$ is large; for example, $\mu(U)\cdot N = \mathcal{O}(N)$ as $N\to\infty$. In this case, (1.2) 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*.

1.2 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 according to an appropriate variable density. This was recently confirmed by mathematical analysis in [3, 4]. The key to this work is to replace the three principles of compressed sensing with three new concepts – *sparsity in levels, multi-level sampling* and *local coherence* – and prove recovery estimates akin to (1.2) under these more general settings.

[1]Here and elsewhere in this section we shall use the notation $a \gtrsim b$ to mean that there exists a constant $C > 0$ independent of all relevant parameters such that $a \geq Cb$.

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

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

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

Definition 1.2 (Multi-level sampling scheme). *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \le N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \le N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that*

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

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

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

as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.

Definition 1.3 (Local coherence). *Let U be an isometry of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. If $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \le N_1 < \ldots N_r$ and $1 \le M_1 < \ldots < M_r$ the $(k,l)^{\text{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ is given by*

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)}, \qquad k, l = 1, \ldots, r, \tag{1.3}$$

where $N_0 = M_0 = 0$ and P_b^a denotes the projection matrix corresponding to indices $\{a+1, \ldots, b\}$.

In [3] a new theory of compressed sensing was introduced based on these new assumptions. Therein, instead of a standard compressed sensing estimate (1.2) determining the total number of measurements, one has the following estimate regarding the local number of measurements m_k in the k^{th} level:

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l \right) \cdot \log(N), \quad k = 1, \ldots, r. \tag{1.4}$$

In particular, the sampling strategy (i.e. the parameters $\mathbf{N}$ and $\mathbf{m}$) is now determined through the local sparsities and incoherences.

1.3 Asymptotic Incoherence

This estimate begs the following question: how do the local sparsity and incoherences behave in practice? As described in [3,4], natural images possess not just sparsity, but so-called asymptotic sparsity. That is, the ratios $s_k/(N_k - N_{k-1}) \to 0$ as $k \to \infty$ in any appropriate basis (e.g. wavelets and their generalizations). Furthermore, such problems are also *asymptotically incoherent*:

Definition 1.4. *Let $U : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ be bounded and linear. Then we say U is 'asymptotically incoherent' if, as $N \to \infty$,*

$$\mu(P_N^{\perp} U), \ \mu(U P_N^{\perp}) \longrightarrow 0. \tag{1.5}$$

Note that the local coherence can be estimated as follows:

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) \le \sqrt{\min\left(\mu(P_{N_{k-1}}^{\perp} U), \mu(U P_{M_{l-1}}^{\perp})\right) \cdot \mu(P_{N_{k-1}}^{\perp} U)}. \tag{1.6}$$

Hence, the combination of asymptotic sparsity and asymptotic incoherence allow the coherence barrier to be broken. Images can be recovered from small numbers of measurements, and the appropriate multilevel sampling strategy is determined via (1.4).

Nevertheless, it is clear from (1.6) that in order to determine the appropriate sampling density one needs good estimates for $\mu(P_N^{\perp} U)$ and $\mu(U P_N^{\perp})$ for $N \in \mathbb{N}$. This is the key contribution of this paper. Our main results provide estimates for the precise convergence rate in (1.5).

1.4 Main Results

In this paper we focus on studying operators $U \in \mathcal{B}(l^2(\mathbb{N}))$ of the form $U_{m,n} = \langle \tau(n), \rho(m) \rangle$ where $(\rho(n))_{n\in\mathbb{N}}, (\tau(n))_{n\in\mathbb{N}}$ enumerates two bases B_1, B_2 of subspaces in a Hilbert space and $\langle \cdot, \cdot \rangle$ denotes the inner product. Our aim is to determine the optimal decay rate of the coherence $\mu(P_N^\perp U)$ and, by swapping the two bases B_1, B_2 around, of $\mu(UP_N^\perp)$. The decay rate of $\mu(P_N^\perp U)$ depends not only on the bases B_1, B_2, but also on the way we enumerate the basis B_1, which could be used to guide, for example, how we might subsample from B_1. This leads us to the following two key questions:

1. What are the fastest possible decay rates for $\mu(P_N^\perp U)$ and $\mu(UP_N^\perp)$?
2. What are examples of orderings of B_1 and B_2 that produce these optimal decay rates?

One important issue is how we interpret these questions; to what degree of accuracy do we want to describe the optimal decay rate and the optimal ordering(s)? It turns out (see Figure 2) that, in the Fourier-wavelet case, simply looking for an ordering with fastest decay can lead to unnecessarily complex orderings that are wavelet dependent. However, if we only want to find the optimal decay rate up to multiplication by a constant we admit optimal orderings that are simple and wavelet independent.

In the general case, i.e. for any pair of bases (B_1, B_2), we have the following result on the fastest possible decay:

Theorem 1.5. *Suppose $U \in \mathcal{B}(l^2(\mathbb{N}))$ is an isometry[2]. Then we cannot, for any $\alpha > 1$, have the decay*

$$\mu(P_N^\perp U) = \mathcal{O}\left(\frac{1}{N^\alpha}\right), \qquad N \to \infty.$$

Futhermore the sum $\sum_N \mu(P_N^\perp U)$ must diverge.

This theorem is directly implied by Theorem 2.14. Furthermore, Lemma 2.16 shows that this statement on the fastest decay of $\mu(P_N^\perp U)$ cannot be strengthened.

In this paper we shall answer questions 1. and 2. for the following specific cases:

Theorem 1.6. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be defined as above. If B_1 is a one-dimensional Fourier basis and B_2 is a one-dimensional Daubechies[3] wavelet basis, then the fastest decay possible is*

$$\mu(P_N^\perp U),\ \mu(UP_N^\perp) = \mathcal{O}\left(\frac{1}{N}\right), \qquad N \to \infty,$$

and this decay can be realised by orderings of the two bases (B_1, B_2). Moreover, by Theorem 1.5, there is no other pair of orthonormal bases (B_1, B_2), with corresponding $U \in \mathcal{B}(l^2(\mathbb{N}))$ an isometry, that can yield faster decay on the asymptotic incoherence as a power of N. Consequently we say that the wavelet with Fourier case has the fastest decaying asymptotic incoherence of any pair of orthonormal bases up to powers of N.

This theorem is an easy to state, however weaker version of Theorem 3.8 which describes this result in full detail. We mention here that the notion of one ordering having a faster decay rate than another is described rigorously in Definition 2.4.

Theorem 1.7. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be defined as above. If B_1 is a one-dimensional Fourier basis and B_2 is a one-dimensional Legendre polynomial basis, then the fastest decay possible is*

$$\mu(P_N^\perp U),\ \mu(UP_N^\perp) = \mathcal{O}\left(\frac{1}{N^{2/3}}\right), \qquad N \to \infty,$$

and this decay can be realised by orderings of the two bases (B_1, B_2).

[2]Which includes the case where B_1, B_2 are orthonormal and the span of B_1 contains B_2.

[3]We include Haar wavelets as a special case.

As with the previous result, this theorem is an easy to state, however weaker version of Theorem 4.4, where we go into further detail.

For these one-dimensional cases we find that canonical orderings give the optimal decay rate; we order the wavelet basis according to their levels, use the natural ordering of the Legendre polynomial basis and order the Fourier basis according to size of frequency.

It should also be mentioned that in all cases we first provide bounds on the row incoherences[4], namely $\mu(\pi_N U)$ and $\mu(U\pi_N)$, where π_N denotes the projection onto the Nth coordinate (see (2.2)). From these bounds on the row incoherences we derive the results given above. For example the Fourier-Wavelet result is originally stated as

$$\mu(\pi_N U), \ \mu(U\pi_N) = \mathcal{O}\left(\frac{1}{N}\right), \qquad N \to \infty.$$

1.5 Outline for the Remainder of the Paper

Before going into more theory, we shall outline the structure for the rest of the paper. We present the general framework in Section 2 where the notions of optimal decay rates and optimal orderings are defined and a few general results for any pair of bases are proved.

After this the analysis is split into cases; (i) the one-dimensional case and (ii) the multi-dimensional case, which is more technical but a natural extension of (i), and is therefore deferred to a later paper. The framework setup in Section 2 is central to the multi-dimensional case as well.

Next we tackle the main theorems presented in the introduction, first considering the one-dimensional Fourier-Wavelet cases in Section 3. We then study the one-dimensional Fourier-polynomial cases in Section 4.

Finally we analyse how the different structures of the Fourier-wavelet and Fourier-polynomial cases lead to differing optimal subsampling schemes with a simple numerical example in Section 5.

2 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,

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

We call (B_1, B_2) a 'basis pair'.

Definition 2.1 (Ordering). *Let S be a set. Say that a function $\rho : \mathbb{N} \to S$ is an 'ordering' of S if it is bijective.*

Definition 2.2 (Change of Basis Matrix). *For a basis pair (B_1, B_2), with corresponding orderings $\rho : \mathbb{N} \to B_1$ and $\tau : \mathbb{N} \to B_2$, form a matrix U by the equation*

$$U_{m,n} := \langle \tau(n), \rho(m) \rangle. \tag{2.1}$$

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

2.1 Comparing Orderings and Decay Rates

We start this section by defining two useful projections.

Definition 2.3. *We define the following linear projection operators from $l^2(\mathbb{N})$ to itself as follows:*

$$Q_N(x)_i := \begin{cases} 0 & i < N \\ x_i & i \geq N \end{cases}, \qquad \pi_N(x)_i := \begin{cases} 0 & i \neq N \\ x_i & i = N \end{cases}. \tag{2.2}$$

[4]Row incoherences have also been studied from a finite-dimensional viewpoint in [19], where they are called "local coherences".

We are interested in studying the asymptotic incoherence of the matrix U, namely we would like to see how $\mu(P_N^{\perp}U)$ behaves as N gets large. For notational simplicity we shall work with $\mu(Q_N U)$ (which is equal to $\mu(P_{N-1}^{\perp}U)$ for $N \geq 2$). We will also be looking at the row coherence of U, defined as $\mu(\pi_N U)$. Notice that if we permute the columns of U then this does not effect $\mu(Q_N U)$ or $\mu(\pi_N U)$, which means that $\mu(Q_N U)$ and $\mu(\pi_N U)$ are independent of the ordering of B_2. Since we want to see how asymptotic incoherence behaves with different orderings, we need a precise way of saying one ordering has a slower decay rate than another:

Definition 2.4 (Relations on the set of orderings). *Let $\rho_1, \rho_2 : \mathbb{N} \to B_1$ be any two orderings of a basis B_1 and τ any ordering of a basis B_2. Let $U^1 := [(B_1, \rho_1), (B_2, \tau)]$, $U^2 := [(B_1, \rho_2), (B_2, \tau)]$ as in (2.1). If there is a constant $C > 0$ such that*

$$\mu(Q_N U^1) \leq C \cdot \mu(Q_N U^2), \qquad \forall N \in \mathbb{N},$$

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

$$[a] \prec [b] \quad \Leftrightarrow \quad a \prec b,$$

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

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

Definition 2.5 (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 shortly in Lemma 2.9 that optimal orderings always exist . Notice that ρ is an optimal ordering if and only if for every other ordering ρ' we have $\rho \prec \rho'$. In order to study the asymptotic incoherence of the matrix U, we need the following definitions.

Definition 2.6. *Let $f, g : \mathbb{N} \to \mathbb{R}_{>0}$. We write $f \lesssim g$ to mean there is a constant $C > 0$ such that*

$$f(N) \leq C \cdot g(N), \qquad \forall N \in \mathbb{N}.$$

If both $f \lesssim g$ and $g \lesssim f$ holds, we write '$f \approx g$'.

Definition 2.7 (Optimal decay rate). *Suppose $\rho : \mathbb{N} \to B_1$ is an optimal ordering for the basis pair (B_1, B_2) and $U = [(B_1, \rho), (B_2, \tau)]$ a corresponding incoherence matrix (with some ordering τ of B_2). Then any decreasing function $f : \mathbb{N} \to \mathbb{R}_{>0}$ which satisfies $f \approx g$, where g is defined by $g(N) = \mu(Q_N U)$, $\forall N \in \mathbb{N}$, is said to be an 'optimal decay rate' of the basis pair (B_1, B_2).*

Notice that an optimal decay rate is unique up to the equivalence relation $\approx$ defined on the set of functions $f : \mathbb{N} \to \mathbb{R}_{>0}$. Next we define the concept of best ordering.

Definition 2.8 (Best ordering). *Let (B_1, B_2) be a basis pair. Then any ordering $\rho : \mathbb{N} \to B_1$ 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 $\mu(\pi_N U) = \mu(Q_N U)$. If ρ' is any other ordering and $U' = [(B_1, \rho'), (B_2, \tau)]$ then since $Q_N U'$ must contain one of the first N rows of U we must have that

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

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

Lemma 2.9. *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).

In either case, optimal orderings always exist.

Proof. Let $\rho : \mathbb{N} \to B_1$, $\tau : \mathbb{N} \to B_2$ 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 $D \subset \mathbb{N}$

$$\sup_{N \in \mathbb{N} \setminus D} \mu(\pi_N U), \tag{2.3}$$

is attained for some $N \in \mathbb{N} \setminus D$. In this case we can then construct a best ordering $\rho^* : \mathbb{N} \to B_1$ 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

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

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 (2.3) 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

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

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

Lemma 2.10. *Suppose that we have a basis pair (B_1, B_2) with two orderings $\rho : \mathbb{N} \to B_1$, $\tau : \mathbb{N} \to B_2$ of B_1, B_2 respectively. If $U = [(B_1, \rho), (B_2, \tau)]$ satisfies*

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

then a best ordering exists.

Proof. The supremum (2.3) is always attained and therefore we fall into case 1) of the previous lemma. ☐

Since $\mu(P_N^{\perp} U) \to 0$ as $N \to \infty$ implies $\mu(\pi_N U) \to 0$ as $N \to \infty$ we also deduce that if U is asymptotically incoherent then there must be a best ordering.

Remark 2.1 Given a basis pair (B_1, B_2), it is tempting just to search for a best ordering. This, however, can be problematic if we want to find optimal orderings that are simple to describe and independent from any factors that are also independent of the optimal decay rate. Figure 2 shows an example where the best orderings are wavelet dependent, even though the optimal equivalence classes can be described in a wavelet independent manner. Although the difference between the best orderings is very minor in Figure 2, this difference becomes more noticeable when working in higher dimensions.

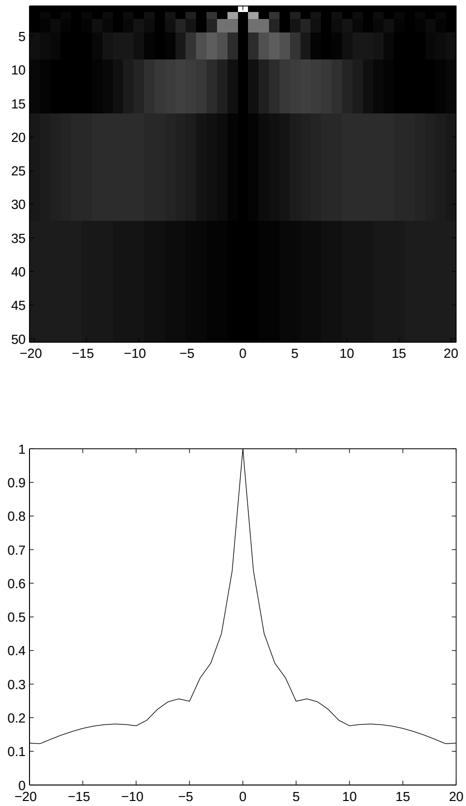
In later sections of this paper we shall first deduce bounds on the row incoherence $\mu(\pi_N U)$ and then use these to deduce bounds on $\mu(Q_N U)$. The next important Lemma describes a few connections between these two notions of coherence.

Lemma 2.11. ***1):*** *Let (B_1, B_2) be a basis pair and τ any ordering of B_2. Furthermore, let $B_1' \subset B_1$ have an ordering $\rho_1 : \mathbb{N} \to B_1'$, and define $U_1 := [(B_1', \rho_1), (B_2, \tau)]$. Suppose that that there is a decreasing function $f_1 : \mathbb{N} \to \mathbb{R}_{>0}$ such that*

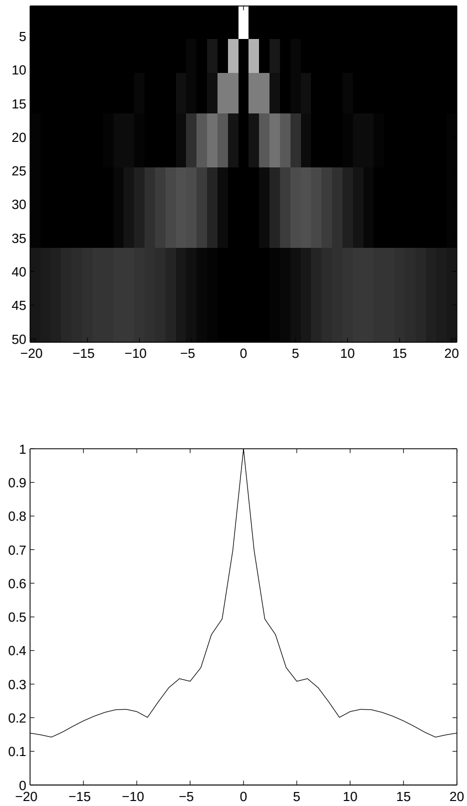
$$f_1(N) \leq \mu(\pi_N U_1), \qquad \forall N \in \mathbb{N}.$$

Then if $\rho_2 : \mathbb{N} \to B_1$ is an ordering, $U_2 = [(B_1, \rho_2), (B_2, \tau)]$ and $f_2 : \mathbb{N} \to \mathbb{R}_{>0}$ is a function with

$$\mu(Q_N U_2) \leq f_2(N), \qquad \forall N \in \mathbb{N},$$



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



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

Figure 2: 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.

then $f_1(N) \leq f_2(N)$ for every $N \in \mathbb{N}$.

2): *Let ρ be an ordering of B_1 with $U := [(B_1, \rho), (B_2, \tau)]$ and $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a decreasing function with $f(N) \to 0$ as $N \to \infty$. If, for some constants $C_1, C_2 > 0$, we have*

$$C_1 f(N) \leq \mu(\pi_N U) \leq C_2 f(N), \qquad \forall N \in \mathbb{N}, \tag{2.4}$$

then ρ is an optimal ordering and f is a representative of the optimal decay rate.

Proof. **1):** Let $\theta(N)$ denote the smallest $m \in \mathbb{N}$ such that $\rho_1(m) \in \{\rho_2(k)\}_{k=N}^{\infty}$ and $m' = m'(N) \in \{N, N+1, \ldots\}$ 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_1 is decreasing we see that

$$f_1(N) \leq f_1(\theta(N)) \leq \mu(\pi_{\theta(N)} U_1) = \mu(\pi_{m'(N)} U_2) \leq \mu(Q_N U_2) \leq f_2(N).$$

2): By (2.4) and Lemma 2.9 we know there is a best ordering ρ^* of B_1 with corresponding matrix $U^* = [(B_1, \rho^*), (B_2, \tau)]$. Furthermore if g is a representative of the optimal decay rate we know that there are constants $D_1, D_2 > 0$ such that for every $N \in \mathbb{N}$,

$$D_1 \cdot g(N) \leq \mu(\pi_N U^*) = \mu(Q_N U^*) \leq D_2 \cdot g(N). \tag{2.5}$$

However, we also know from f being decreasing and (2.4) that

$$C_1 \cdot f(N) \leq \mu(\pi_N U) \leq \mu(Q_N U) = \max_{N' \geq N} \mu(\pi_{N'} U) \leq C_2 \cdot \max_{N' \geq N} f(N') = C_2 \cdot f(N). \tag{2.6}$$

Therefore we may apply part 1 of the Lemma twice to (2.5) & (2.6) to deduce $f \lesssim g$ and $g \lesssim f$, which implies $f \approx g$. $\square$

Definition 2.12 (Strongly optimal ordering)**.** *Let (B_1, B_2) be a basis pair and τ any ordering of B_2. Then any ordering ρ of B_1 that satisfies (2.4) for some decreasing function f is said to be a 'strongly optimal ordering'.*

Notice that any best ordering is also a strongly optimal ordering. Furthermore if g is a representative of the optimal decay rate then by Lemma 2.11 an ordering ρ is strongly optimal if and only if there are constants $C_1, C_2 > 0$ such that

$$C_1 \cdot g(N) \leq \mu(\pi_N U) \leq C_2 \cdot g(N), \qquad \forall N \in \mathbb{N}.$$

Throughout this paper we would like to define an ordering according to a particular property of the 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 2.13 (Consistent ordering)**.** *Let $F : S \to \mathbb{R}$ where S is a set. We say that an ordering $\rho : \mathbb{N} \to S$ is 'consistent with respect to F' if*

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

Before moving onto specific cases, we consider the general case where U is an isometry and ask; is there a universal lower bound on the incoherence?

Theorem 2.14. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then $\sum_N \mu(Q_N U)$ diverges.*

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

$$\sum_{N=N'}^{\infty} |u_{N,j}|^2 \leq \sum_{N=N'}^{\infty} \mu(Q_N U) \leq 1/4^2, \qquad j \in \mathbb{N} \tag{2.7}$$

Now define the vectors

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

Inequality (2.7) 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

$$\sup_{\substack{j,j' \in \{1,\dots,M\} \\ j \neq j'}} |\langle w_j, w_{j'} \rangle| \to 1, \quad \text{as} \quad M \to \infty. \tag{2.8}$$

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 $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \varnothing$, 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 subcover[5], contradicting compactness of the unit ball in $\mathbb{C}^{N'-1}$. Since $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \varnothing,\ w_{j_1} \in B_\delta(w_{j_2}) \subset W_{j_2}(\epsilon)$ and so $|\langle w_{j_1}, w_{j_2} \rangle| > 1 - \epsilon$. Since $\epsilon > 0$ was arbitrary we have proved (2.8).

Therefore, by (2.8) 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

$$|\langle v_{j_1}^1, v_{j_2}^1 \rangle| > \frac{1}{2} \|v_{j_1}^1\|_2 \|v_{j_2}^1\|_2 > \frac{3^2}{2 \cdot 4^2}. \tag{2.9}$$

[5]Any finite subcover would miss infinitely many of the points w_j.

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

$$|\langle v_{j_1}^2, v_{j_2}^2\rangle| \le \|v_{j_1}^2\|_2 \|v_{j_1}^2\|_2 \le \frac{1}{4^2}. \tag{2.10}$$

Therefore, combining (2.9) with (2.10) gives us

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

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. ☐

Corollary 2.15. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be any isometry. Then there does not exist an $\epsilon > 0$ such that*

$$\mu(Q_N U) = \mathcal{O}(N^{-1-\epsilon}), \qquad N \to \infty.$$

Noting the above corollary and that $\mu(U) = N^{-1}$ is the best result possible for the finite $U \in \mathbb{C}^N \times \mathbb{C}^N$ case, it might be tempting to believe $\mu(Q_N U) = \mathcal{O}(N^{-1})$ is the best decay rate we can achieve for an isometry. However, it turns out that Theorem 2.14 cannot be improved without imposing additional conditions on U:

Lemma 2.16. *Let $f, g : \mathbb{N} \to \mathbb{R}$ 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*

$$\mu(Q_N U) \le f(N), \quad \mu(U Q_N) \le g(N), \qquad N \in \mathbb{N}. \tag{2.11}$$

Proof. The proof is constructive. We may assume without loss of generality that $f(N),\ g(N) \le 1$ for all $N \in \mathbb{N}$. We will construct a matrix $U = (u_{i,j})_{i,j\in\mathbb{N}}$ satisfying (2.11) 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 (for[6] $N \in \mathbb{N}$)

$$(v_j)_N = \begin{cases} \big(g(j)f(N)\big)^{1/2}, & \text{if } \sum_{i=1}^{N-1}((v_j)_i)^2 + g(j)f(N) \le 1, \quad N \in \Omega_j, \\ \big(1 - \sum_{i=1}^{N-1}((v_j)_i)^2\big)^{1/2}, & \text{if } \sum_{i=1}^{N-1}((v_j)_i)^2 \le 1, \\ & \quad\ \sum_{i=1}^{N-1}((v_j)_i)^2 + g(j)f(N) \ge 1, \quad N \in \Omega_j, \\ 0, & \text{Otherwise.} \end{cases} \tag{2.12}$$

It is immediate from the definition that v_j is supported on Ω_j and $((v_j)_N)^2 \le f(N)g(j)$ for every $N, j \in \mathbb{N}$ which implies that (2.11) holds. Furthermore, it easy to show by induction on N that $\|v_j\|_2 \le 1$. 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

$$\sum_{\substack{N\in\Omega_j \\ N\le N'}} g(j)f(N) \ge 1, \qquad \sum_{\substack{N\in\Omega_j \\ N\le N'-1}} g(j)f(N) \le 1.$$

For $N \le N' - 1, N \in \Omega_j$ we fall into the first case of (2.12), 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. ☐

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 (2.11) holds for an isometry U is $\alpha = 1$, which shall be attained by our first example.

From hereon in we shall work with specific bases and find optimal orderings and decay rates for each case.

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

3 1D Fourier-Wavelet Case

In this section we shall be using a Fourier basis and a wavelet basis. However we shall consider two types of wavelet basis in this section: standard wavelets and boundary wavelets. Before doing this this, we will first define and order the Fourier basis.

Let $\epsilon > 0$ be fixed. For $x \in \mathbb{R}$, define

$$\chi_k(x) \ = \ \sqrt{\epsilon}\exp(2\pi \mathrm{i}\epsilon k x) \ \cdot \ \mathbb{1}_{[(-2\epsilon)^{-1},(2\epsilon)^{-1}]}(x), \qquad k \in \mathbb{Z}. \tag{3.1}$$

Notice that $(\chi_k)_{k\in\mathbb{Z}}$ is a basis for $L^2[(-2\epsilon)^{-1},(2\epsilon)^{-1}]$. We set $B_\mathrm{f} = B_\mathrm{f}(\epsilon) := (\chi_k)_{k\in\mathbb{Z}}$. (The little f here stands for 'Fourier').

Definition 3.1 (Standard ordering)**.** *We define $F_\mathrm{f} : B_\mathrm{f} \to \mathbb{N}\cup\{0\}$ by $F_\mathrm{f}(\chi_k) = |k|$ and say that an ordering $\rho : \mathbb{N} \to B_\mathrm{f}$ is a 'standard ordering' if it is consistent with F_f (recall Definition 2.13).*

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

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

Definition 3.1 says that an ordering ρ of $B_\mathrm{f}(\epsilon)$ is standard if and only if the function $|\lambda\circ\rho|$ is nondecreasing. Therefore ρ is standard 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 if ρ is standard then $|\lambda\circ\rho(m)| = \lceil(m-1)/2\rceil$. We now define and order the two types of wavelet basis.

3.1 Case 1 - Standard Wavelets

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

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

We write

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$
$$V_j := \overline{\mathrm{Span}\{\phi_{j,k} : k \in \mathbb{Z}\}}, \quad W_j := \overline{\mathrm{Span}\{\psi_{j,k} : k \in \mathbb{Z}\}}.$$

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

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

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 set[7]

$$B_\mathrm{w} := \left\{ \ \phi_{J,k}, \ \psi_{j,k} : \begin{array}{c} \mathrm{Supp}(\phi_{J,k}) \cap (-1,1) \neq \emptyset, \\ \mathrm{Supp}(\psi_{j,k}) \cap (-1,1) \neq \emptyset, \\ j \in \mathbb{N}, j \geq J, \ k \in \mathbb{Z} \end{array} \right\}, \tag{3.2}$$

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 $\mathrm{supp}(f) \subseteq [-1,1]$ 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 3.2 (Leveled ordering (standard wavelets))**.** *Define $F_\mathrm{w} : B_\mathrm{w} \to \mathbb{R}$ by*

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

and say that any ordering $\tau : \mathbb{N} \to B_\mathrm{w}$ is a 'leveled ordering' if it is consistent with F_w.

Notice that $F_\mathrm{w}(\psi_{j,k}) = j$. 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.

[7] 'w' here stands for 'wavelet'.

3.2 Case 2 - 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 [24, 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 $\mathrm{Supp}(\psi) = \mathrm{Supp}(\phi) = [-p+1, p]$, boundary scaling functions and wavelets (using the same notation as in [24] [8])

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

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

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

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

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

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

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

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

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

3.3 Proof of Theorem 1.6

Suppose that ρ is an ordering of the Fourier basis $B_{\text{f}} = B_{\text{f}}(\epsilon)$, τ is an ordering of a wavelet basis B_{w} (or B_{bw}) and set $U = [(B_{\text{f}}(\epsilon), \rho), (B_{\text{w}}, \tau)]$ (or $U = [(B_{\text{f}}(\epsilon), \rho), (B_{\text{bw}}, \tau)]$). Recall that the basis $B_{\text{f}}(\epsilon)$ spans $L^2[(-2\epsilon)^{-1}, (2\epsilon)^{-1}]$.

Remark 3.1 For standard wavelets 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 $B_{\text{f}}(\epsilon)$. 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.

With U defined as above (with either B_{w} or B_{bw}), the key observations for handling the entries of U are

$$\begin{aligned} U_{m,n} = \langle \tau(n), \rho(m) \rangle &= \int_{\mathbb{R}} \sqrt{\epsilon} \exp(-2\pi i \epsilon x \cdot \lambda \circ \rho(m)) \cdot \tau(n)(x)\, dx \\ &= \sqrt{\epsilon} \mathcal{F}\tau(n)(\epsilon \cdot \lambda \circ \rho(m)), \end{aligned} \tag{3.3}$$

recalling that $\mathcal{F}$ denotes the Fourier Transform. We also observe that

$$\begin{aligned} &\mathcal{F}\phi_{j,k}(\omega) = e^{-2\pi \mathrm{i} 2^{-j}k\omega} 2^{-j/2} \mathcal{F}\phi(2^{-j}\omega), \qquad \mathcal{F}\psi_{j,k}(\omega) = e^{-2\pi \mathrm{i} 2^{-j}k\omega} 2^{-j/2} \mathcal{F}\psi(2^{-j}\omega), \\ &\mathcal{F}\psi_{j,n}^{\text{left}}(\omega) = 2^{-j/2} e^{2\pi \mathrm{i}} \mathcal{F}\psi_n^{\text{left}}(2^{-j}\omega), \qquad \mathcal{F}\psi_{j,n}^{\text{right}}(\omega) = 2^{-j/2} e^{-2\pi \mathrm{i}} \mathcal{F}\psi_n^{\text{right}}(2^{-j}\omega). \end{aligned} \tag{3.4}$$

We now come to our first optimal incoherence estimate.

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

Proposition 3.4. *Let τ be any leveled ordering of a standard wavelet basis and $U = [(B_{\mathrm{w}}, \tau), (B_{\mathrm{f}}(\epsilon), \rho)]$ for any ordering ρ of the Fourier Basis $B_{\mathrm{f}}(\epsilon)$. 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*

$$\frac{\epsilon \cdot C_1}{N} \leq \mu(\pi_N U) \leq \frac{\epsilon \cdot C_2}{N}. \tag{3.5}$$

Furthermore, suppose that τ is instead an ordering of a boundary wavelet basis and $U = [(B_{\mathrm{bw}}, \tau), (B_{\mathrm{f}}(\epsilon), \rho)]$ for any ordering ρ of the Fourier Basis $B_{\mathrm{f}}(\epsilon)$. 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 $N \in \mathbb{N}$, (3.5) holds.

Proof. By equation (3.3) we know that (since $\lambda \circ \rho : \mathbb{N} \to \mathbb{Z}$ 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 $j(N) := F_{\mathrm{w}}(\tau(N))$ 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 (3.4),

$$\mu(\pi_N U) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-J} |\mathcal{F}\phi(2^{-J}\epsilon m)|^2. \tag{3.6}$$

Furthermore, for the wavelet terms in B_{w}, which correspond to $N \geq 2^{J+1} + a$, we have that, by (3.4),

$$\mu(\pi_N U) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-j(N)} |\mathcal{F}\psi(2^{-j(N)}\epsilon m)|^2. \tag{3.7}$$

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 this 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 $j(N) \to \infty$ as $N \to \infty$ because the ordering τ is leveled, we find that

$$\frac{\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(\epsilon 2^{-j(N)} m)|^2}{\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2} \longrightarrow 1 \quad \text{as} \quad N \to \infty. \tag{3.8}$$

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

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 $f \in B_{\mathrm{w}}, F_{\mathrm{w}}(f) = j(N) - 1$ 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 we have the inequality, for $j(N) \geq J$,

$$2^{j(N)} \leq N \leq 2^{j(N)+a+2}. \tag{3.9}$$

Now we will tackle the upper and lower bounds of (3.5) separately:

Upper Bound: We will show that $\mu(\pi_N U) \leq \frac{\epsilon \cdot C_2}{N}$. Notice from (3.9) 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 (3.7) 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: We will show that $\frac{\epsilon \cdot C_1}{N} \leq \mu(\pi_N U)$. First notice that from (3.9) that we have the lower bound $2^{-j(N)} \geq N^{-1}$. Next notice that from (3.8) 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

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

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 $N > 2^{J+1}$,

$$2^{j(N)} \leq N \leq 2^{j(N)+2}. \tag{3.10}$$

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. □

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

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \tag{3.11}$$

This condition holds for all Daubechies wavelets (see the proof of Proposition 4.7 in [11]), in fact it even holds if we change the power of ω from $1/2$ to 1.

Lemma 3.5. *Let ϕ be a Daubechies scaling function, with corresponding mother wavelet ψ. Then, along with (3.11), we also have*

$$|\mathcal{F}\psi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \tag{3.12}$$

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

$$|\mathcal{F}\phi_n^{\text{left}}(\omega)|, \ |\mathcal{F}\phi_n^{\text{right}}(\omega)|, \ |\mathcal{F}\psi_n^{\text{left}}(\omega)|, \ |\mathcal{F}\psi_n^{\text{right}}(\omega)| \leq \frac{K}{|\omega|^{1/2}}, \tag{3.13}$$

along with (3.11) and (3.12). In fact (3.12) and (3.13) hold with the powers of $1/2$ replaced by 1.

Proof. We notice that if (3.11) holds then we can use the equation (see (2.14) in [18])

$$\mathcal{F}\psi(2\omega) = \exp(2i\pi\omega) \cdot \nu(2\omega) \cdot m_0(\omega + 1/2) \cdot \mathcal{F}\phi(\omega), \tag{3.14}$$

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 1[9]. 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 [18]) we can show that (3.11) 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 (3.11) is closed under finite linear combinations. Next observe that if we prove an estimate of the form (3.11) for the functions (see page 71 of [10])

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

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 $\mathcal{F}T_n(\omega) = \big(\mathcal{F}\phi_{0,-n+p-1} * \mathcal{F}\mathbb{1}_{[0,2p-1]}\big)(\omega)$. Now we make two observations:

1. $|\mathcal{F}\mathbb{1}_{[0,2p-1]}(\omega)| = |(\exp(-2\pi \mathrm{i}(2p-1)\omega)-1)\cdot(2\pi\mathrm{i}\omega)^{-1}| \leq C_1 \cdot (|\omega|+1)^{-1}$ 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 [11]).

We now that claim that if two functions f, g satisfy

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

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 $\omega > 0$)

$$\begin{aligned} |f * g(\omega)| \cdot |\omega| &\leq C_1 C_2 \int_{\mathbb{R}} \frac{|\omega|}{(|u|+1)(|\omega-u|+1)^{1+\alpha}}\, du \\ &\leq C_1 C_2 \Bigg(\int_{-\infty}^{\omega/2} \frac{|\omega|}{(|u|+1)(|\omega-u|+1)^{1+\alpha}}\, du \\ &\qquad + \int_{\omega/2}^{+\infty} \frac{|\omega|}{(|u|+1)(|\omega-u|+1)^{1+\alpha}}\, du \Bigg), \end{aligned} \tag{3.15}$$

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

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

To bound the last integral in (3.15) 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. □

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

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

Lemma 3.6. *For any compactly supported wavelet ψ with a 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 (3.14) that

$$|\mathcal{F}\psi(2\omega)| \ = \ |m_0(\omega + 1/2)| \cdot |\mathcal{F}\phi(\omega)|. \tag{3.16}$$

Furthermore, we also know that $|\mathcal{F}\phi(0)| = 1$ and $m_0(1/2) = 0$ [18][10]. However, since ϕ is compactly supported, m_0 is a non-zero trigonometric polynomial so it follows that this zero at $1/2$ is isolated. Therefore, since $\mathcal{F}\phi$ is continuous, we deduce that (3.16) is nonzero when $\omega > 0$ is sufficiently small. □

We now prove an optimal incoherence result for $U_1 = [(B_{\mathrm{f}}(\epsilon), \rho), (B_{\mathrm{w}}, \tau)]$. Notice that this matrix is different to $U_2 = [(B_{\mathrm{w}}, \tau), (B_{\mathrm{f}}(\epsilon), \rho)]$ which was covered in Proposition 3.4. In particular, $U_1 = (U_2)^*$ (the adjoint of U_2) and so have that $\mu(\pi_N U_1) = \mu(U_2 \pi_N)$.

Proposition 3.7. *Let ρ be any standard ordering of the Fourier basis $B_{\mathrm{f}}(\epsilon)$ and $U = [(B_{\mathrm{f}}(\epsilon), \rho), (B_{\mathrm{w}}, \tau)]$ for any standard wavelet ordering τ. 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 $\epsilon \in (0, 1/2]$.

Proof. **Upper Bound:** Since ρ is a standard ordering if $m = 1$ then $\lambda \circ \rho(m) = 0$ and since $|\mathcal{F}\phi(0)| = 1$, $\mathcal{F}\psi(0) = 0$ (see (3.16) 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 (3.11) 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 standard 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 (3.11) 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.

[10] See Section 2 Theorem 1.7 and Equation (3.1) in the reference.

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 3.6, 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 3.6 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 3.2 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

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

Summarising the consequences of what we have proved in this section, while throwing away ϵ dependence from our results, we have the following theorem.

Theorem 3.8. *Let the Fourier basis $B_{\mathrm{f}}(\epsilon)$ be defined as in (3.1) and a wavelet basis B_{w} be defined as in Section 3.1 with $\epsilon \in I_{J,p}$. Let ρ be a standard ordering of $B_{\mathrm{f}}(\epsilon)$, τ a leveled ordering of B_{w} and $U = [(B_{\mathrm{f}}(\epsilon), \rho), (B_{\mathrm{w}}, \tau)]$. Furthermore, suppose that (3.11) holds for the wavelet basis and ϵ is kept fixed. Then we have, for some constants $C_1, C_2 > 0$ the decay*

$$\frac{C_1}{N} \leq \mu(P_N^{\perp}U),\ \mu(UP_N^{\perp}) \leq \frac{C_2}{N}, \qquad \forall N \in \mathbb{N}, \tag{3.17}$$

which for either of the coherences $\mu(P_N^{\perp}U),\ \mu(UP_N^{\perp})$, the decay cannot be improved by changing the orderings ρ, τ, except up to alteration of the constants $C_1, C_2 > 0$. Furthermore, the ordering ρ is strongly optimal for the basis pair $(B_{\mathrm{f}}(\epsilon), B_{\mathrm{w}})$ and the ordering τ is strongly optimal for the basis pair $(B_{\mathrm{w}}, B_{\mathrm{f}}(\epsilon))$. Moreover, there is no other pair of bases (B_1, B_2) and orderings ρ, τ, with corresponding $U = [(B_1, \rho), (B_2, \tau)] \in \mathcal{B}(l^2(\mathbb{N}))$ an isometry, that can yield faster decay on the asymptotic incoherence as a power of N; namely we cannot, for any $\alpha > 1$, have the decay

$$\mu(P_N^{\perp}U),\ \mu(UP_N^{\perp}) = \mathcal{O}\left(\frac{1}{N^{\alpha}}\right) \qquad N \to \infty.$$

Finally, if we replace the basis B_{w} with B_{bw}, defined in Section 3.2, in the setup above and also replace the constraint $I_{J,p}$ by $\epsilon \in (0, 1/2]$, the same conclusions also hold.

Proof. Proposition 3.7 gives us the following bound, for some constants $C_1, C_2 > 0$, when $\epsilon \in I_{J,p}$ is fixed[11]:

$$\frac{C_1}{N} \leq \mu(\pi_N U) \leq \frac{C_2}{N}, \qquad N \in \mathbb{N}.$$

Likewise Proposition 3.4 gives us the same bounds for $\mu(U\pi_N)$. Using these bounds we deduce (3.17), with a change of the constants, since

$$\frac{C_1}{N+1} \leq \mu(\pi_{N+1}U) \leq \mu(P_N^{\perp}U) = \max_{N'\geq N+1} \mu(\pi_{N'}U) \leq \max_{N'\geq N+1} \frac{C_2}{N'} = \frac{C_2}{N+1}.$$

The statements about strong optimality follow from Lemma 2.11. Strongly optimality implies optimality, and therefore the statement about not being able to improve the decay rate follows. Finally, the statement about attaining the fastest decay rate for an isometry (up to powers of N) follows from Theorem 2.14. □

[11] When ϵ is fixed we need not worry about the condition $N \geq 1 + 2^{J+1}\epsilon^{-1}$ since we can just change the lower bound constant C_1 to incorporate the finitely many positive terms $\mu(\pi_N U)$ for $N < 1 + 2^{J+1}\epsilon^{-1}$.

4 1D Fourier-Polynomial Case

If $(p_n)_{n\in\mathbb{N}}$ denotes the standard Legendre polynomials on $[-1,1]$ (so $p_n(1)=1$) then the L^2-normalised Legendre polynomials are defined by $\tilde{p}_n = \sqrt{n-1/2}\cdot p_n$ and we write $B_{\mathrm{p}} := (\tilde{p}_n)_{n=1}^{\infty}$ (the p here stands for "polynomial"). B_{p} is already ordered; call this the *natural ordering* . We do not consider any alternative orderings and shall show shortly that the natural ordering is optimal with the Fourier basis.

Before we start covering the incoherences for these two bases we will first need to prove a preliminary result.

Lemma 4.1. *Let J_n denote the nth Bessel function of the first kind and let $j'_{n,k}$ denote the kth non-negative root of J'_n. Furthermore, Let j_n denote the nth spherical Bessel function of the first kind and let $a'_{n,k}$ denote the kth 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 [26, 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)]

$$j_n(x) = \sum_{m=0}^{\infty}\frac{(-1)^m 2^{n+1}(n+m+1)!x^{n+2m}}{m!(2(n+m+1))!}. \tag{4.1}$$

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

$$x^2 j_n(x)'' + 2xj'_n(x) + (x^2 - n(n+1))j_n(x) = 0. \tag{4.2}$$

We first observe that by (4.1), $|j_n(-x)| = |j_n(x)|$, $\forall x\in\mathbb{R}$ and so we need only consider $\sup_{x\in[0,+\infty)}|j_n(x)|$. (4.2) can be rephrased as

$$\big(x^2 j'_n(x)\big)' = (n(n+1)-x^2)j_n(x).$$

Therefore, noting that by (4.1), $j_n(x)>0$ 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

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

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 for $x > n(n+1)$:

$$\begin{aligned}\Lambda'_n(x) &= 2j'_n(x)j_n(x) + \frac{2xj_n'^2(x) + 2x^2 j'_n(x)j''_n(x)}{x^2-n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2-n(n+1))^2} \\ &= \frac{2j'_n(x)\big(j_n(x)(x^2-n(n+1)) + xj'_n(x) + x^2 j''_n(x)\big)}{x^2-n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2-n(n+1))^2} \\ &= -\frac{2xj_n'^2(x)}{x^2-n(n+1)} - \frac{2x^3 j_n'^2(x)}{(x^2-n(n+1))^2} < 0. \qquad \text{(using (4.2))}\end{aligned} \tag{4.3}$$

This tells that $|j_n(a'_{n,1})| > |j_n(a'_{n,2})| > |j_n(a'_{n,3})|....$ To finish the proof we notice that by (4.1), $j_n(0)=0$ for $n\geq 1$ and furthermore, by [1, Eqn. (10.1.14)],

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

and therefore $j_n(x)\to 0$ as $x\to+\infty$ 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. □

Proposition 4.2. *Let τ be the natural ordering of the Legendre polynomial basis and $U = [(B_{\mathrm{p}},\tau),(B_{\mathrm{f}}(\epsilon),\rho)]$ for any ordering ρ of the Fourier basis $B_{\mathrm{f}}(\epsilon)$. Then there are constants $C_1, C_2 > 0$ such that for all $\epsilon\in(0,0.45]$ and $N\in\mathbb{N}$,*

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

Proof. **Upper Bound:** First notice that

$$\begin{aligned} U_{m,n} &= \langle \rho(n), \tilde{p}_m \rangle_{L^2([-1,1])} \\ &= \sqrt{\epsilon} \cdot \sqrt{m-1/2} \int_{-1}^{1} e^{2\pi \mathrm{i} \lambda \circ \rho(n) \epsilon t} p_m(t)\, dt \\ &= \mathrm{i}^{m-1} 2\sqrt{\epsilon(m-1/2)} \cdot j_{m-1}(2\pi\epsilon\lambda \circ \rho(n)) \\ &= \mathrm{i}^{m-1} \frac{\sqrt{m-1/2}}{\sqrt{\lambda \circ \rho(n)}} \cdot J_{m-1/2}(2\pi\epsilon\lambda \circ \rho(n)), \qquad (n \neq 1) \end{aligned} \tag{4.4}$$

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:

$$j_m(z) = \sqrt{\frac{\pi}{2z}} J_{m+1/2}(z). \tag{4.5}$$

Therefore, we find

$$\mu(\pi_N U) \leq 4\epsilon(N-1/2) \sup_{t \in \mathbb{R}} j_{N-1}^2(t). \tag{4.6}$$

We therefore need to estimate $\sup_{t\in\mathbb{R}} |j_m(t)|$. By Lemma 4.1, we know that $\sup_{t\in\mathbb{R}} |j_m(t)| = |j_m(a'_{m,1})|$ for $m \geq 1$, where $a'_{m,1}$ denotes the first positive root of j'_m.

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

$$j_m(a'_{m,1}) \sim \gamma(m+1/2)^{-5/6} + \mathcal{O}((m+1/2)^{-3/2}), \tag{4.7}$$

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 (4.6) we get the upper bound

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

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 (4.6) to cover the case $N \leq N'$, completing the upper bound.

Lower Bound: We focus on the following equation taken from (4.4)

$$|U_{m,n}| = \frac{\sqrt{m-1/2}}{\sqrt{|\lambda \circ \rho(n)|}} \cdot |J_{m-1/2}(2\pi\epsilon\lambda \circ \rho(n))|, \qquad (n \neq 1). \tag{4.8}$$

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

$$j'_{\nu,1} \sim \nu + \zeta\nu^{1/3} + \mathcal{O}(\nu^{-1/3}), \tag{4.9}$$

$$J(j'_{\nu,1}) \sim \kappa \cdot \nu^{-1/3} + \mathcal{O}(\nu^{-1}), \tag{4.10}$$

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

1. $2J'_\nu(x) = J_{\nu-1}(x) - J_{\nu+1}(x)$, [26, p. 45],
2. $\sup_{x\in\mathbb{R}} |J_\nu(x)| = |J_\nu(j'_{\nu,1})|$, using Lemma 4.1 .

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 (4.11) below.

By the triangle inequality $|J_{m-1/2}(k_m)| \geq |J_{m-1/2}(j'_{m-1/2,1})| - |J_{m-1/2}(k_m) - J_{m-1/2}(j'_{m-1/2,1})|$ and we bound the latter term by using integrals[12]:

$$\begin{aligned}
|J_{m-1/2}(k_m) - J_{m-1/2}(j'_{m-1/2,1})| &= \left| \int_{j'_{m-1/2,1}}^{k_m} J'_{m-1/2,1}(t)\, dt \right| \\
&= \left| \int_{j'_{m-1/2,1}}^{k_m} \int_{j'_{m-1/2,1}}^{t} J''_{m-1/2,1}(u)\, du\, dt \right| \\
&\leq \left| \int_{j'_{m-1/2,1}}^{k_m} \int_{j'_{m-1/2,1}}^{t} |J_{m-1/2}(j'_{m-1/2,1})|\, du\, dt \right| \\
&\leq \frac{|j'_{m-1/2,1} - k_m|^2}{2} \cdot |J_{m-1/2}(j'_{m-1/2,1})| \\
&\leq \frac{(\pi\epsilon)^2}{2} \cdot |J_{m-1/2}(j'_{m-1/2,1})|.
\end{aligned} \tag{4.11}$$

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 (4.11) becomes

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

and therefore we deduce

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

Combining this inequality with (4.9), (4.10) gives us the following bound:

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

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

$$\frac{\sqrt{m-1/2}}{\sqrt{k_m}} \cdot |J_{m-1/2}(k_m)| \geq C m^{-1/3}. \tag{4.12}$$

Therefore, given $m \geq M$, let $n(m) \in \mathbb{N}$ be such that $2\pi\epsilon\lambda \circ \rho(n) = k_m$. Then by (4.12) and (4.8) we have

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

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

For $N \leq M$ we observe that from (4.4)

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

As before we observe that since $j_{N-1}(x) \to 0$ as $x \to \infty$, the supremum $\sup_{n\in\mathbb{Z}} |j_{N-1}(2\pi\epsilon n)|$ is a continuous function of ϵ and moreover the supremum converges to $\sup_{x\in\mathbb{R}} |j_{N-1}(x)| > 0$ as $\epsilon \to 0$. 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

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

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

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

Like in Section 3 we now prove an optimal incoherence result for when we swap around the two bases.

Proposition 4.3. *Let ρ be any standard ordering of the Fourier basis $B_{\mathrm{f}}(\epsilon)$ and $U = [(B_{\mathrm{f}}(\epsilon), \rho), (B_{\mathrm{p}}, \tau)]$ for any Legendre polynomial basis ordering τ. 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 (4.4) (with m, n swapped because the U in the theorem is the adjoint of the one in (4.4)) we have

$$|U_{m,n}|^2 = \frac{n - 1/2}{|\lambda \circ \rho(m)|} J^2_{n-1/2}(2\pi\epsilon\lambda \circ \rho(m)) \tag{4.13}$$

$$= 4\epsilon(n - 1/2) j^2_{n-1}(2\pi\epsilon\lambda \circ \rho(m)). \tag{4.14}$$

We shall first derive two useful bounds; notice that if we apply (4.7) to (4.14) then we get the bound, for some constant $\beta > 0$,

$$|U_{m,n}|^2 \leq 4\epsilon(n - 1/2) \cdot (\beta(n - 1/2)^{-5/6})^2 \leq 4\epsilon\beta^2(n - 1/2)^{-2/3}. \tag{4.15}$$

Secondly we shall use the following inequality from [20]

$$|J_\nu(x)| \leq b\nu^{-1/3} \qquad \nu > 0, \quad x \in \mathbb{R}, \tag{4.16}$$

where $b > 0$ is some constant. Applying this to (4.13) gives the bound

$$|U_{m,n}|^2 \leq \frac{n - 1/2}{|\lambda \circ \rho(m)|}(b(n - 1/2)^{-1/3})^2 \leq \frac{b^2(n - 1/2)^{1/3}}{|\lambda \circ \rho(m)|}. \tag{4.17}$$

Recall that our goal is to estimate $|U_{m,n}|$ uniformly in n as $m \to \infty$. We first apply the case $n - 1/2 \geq \epsilon|\lambda \circ \rho(m)|$ to (4.15) 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 (4.17) 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 (4.1)) we deduce that $\mu(\pi_1 U) = \epsilon|j_0(0)|^2 = \epsilon$ which is a stronger bound than required.

Lower Bound: By (4.9) we know that

$$j'_{n+1/2,1} - j'_{n-1/2,1} \to 1 \quad \text{as} \quad n \to \infty. \tag{4.18}$$

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 (4.18) we observe

$$|j'_{n(m)-1/2,1} - |2\pi\epsilon\lambda \circ \rho(m)|| \leq 1/2 + \eta(m, \epsilon), \tag{4.19}$$

where η is such that $\eta(m, \epsilon) \to 0$ as $m \to \infty$ for any fixed ϵ. By using the same method as in (4.11) 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 $\xi(m,\epsilon) \to 8^{-1}$ as $m \to \infty$ with ϵ fixed. This tells us that

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

Combining this with (4.13) we see that, using (4.10),

$$\begin{aligned}|U_{m,n(m)}|^2 &= \frac{n(m)-1/2}{|\lambda\circ\rho(m)|}|J^2_{n(m)-1/2}(2\pi\epsilon\lambda\circ\rho(m))| \\ &\geq \frac{n(m)-1/2}{|\lambda\circ\rho(m)|}\cdot\Big(1-\xi(m,\epsilon)\Big)^2\cdot|J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})|^2 \\ &\geq \frac{n(m)-1/2}{|\lambda\circ\rho(m)|}\cdot\Big(1-\xi(m,\epsilon)\Big)^2\big(\kappa(n(m)-1/2)^{-1/3}+\mathcal{O}((n(m)-1/2)^{-1})\big)^2.\end{aligned} \tag{4.20}$$

By (4.9), (4.19) and the fact that ρ is a standard ordering we know that (for ϵ fixed)

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

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 $\mu(\pi_N U) \geq C\cdot\epsilon^{1/3}\cdot m^{-2/3}$. □

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

Theorem 4.4. *Let the Fourier basis $B_{\mathrm{f}}(\epsilon)$ be defined as in (3.1) and a Legendre polynomial basis B_{p} be defined as in the start of this section with $\epsilon \in (0, 1/2]$. Let ρ be a standard ordering of $B_{\mathrm{f}}(\epsilon)$ and τ a natural ordering of B_{p} and $U = [(B_{\mathrm{f}}(\epsilon), \rho), (B_{\mathrm{p}}, \tau)]$. Then, keeping $\epsilon \in (0, 0.45]$ fixed we have, for some constants $C_1, C_2 > 0$ the decay*

$$\frac{C_1}{N^{2/3}} \leq \mu(P_N^{\perp}U),\ \mu(UP_N^{\perp}), \leq \frac{C_2}{N^{2/3}}, \qquad \forall N \in \mathbb{N}. \tag{4.21}$$

For either of the coherences $\mu(P_N^{\perp}U),\ \mu(UP_N^{\perp})$, the decay cannot be improved by changing the orderings ρ, τ, except up to alteration of the constants $C_1, C_2 > 0$. Furthermore, the ordering ρ is strongly optimal for the basis pair $(B_{\mathrm{f}}(\epsilon), B_{\mathrm{p}})$ and τ is strongly optimal for the basis pair $(B_{\mathrm{p}}, B_{\mathrm{f}}(\epsilon))$.

Proof. We first observe that by Propositions 4.2 & 4.3 with ϵ fixed we have that, for some constants $C_1, C_2 > 0$ and $M \in \mathbb{N}$,

$$\frac{C_1}{N^{2/3}} \leq \mu(\pi_N U),\ \mu(U\pi_N), \leq \frac{C_2}{N^{2/3}}, \qquad \forall N \geq M.$$

This can then be extended to all $N \in \mathbb{N}$, with perhaps a change the constants $C_1, C_2 > 0$, by observing that $\mu(\pi_N U), \mu(U\pi_N)$ for any N are strictly positive (and constant since we have fixed ϵ).

We therefore deduce strong optimality of ρ, τ and inequality (4.21) from Lemma 2.11. □

5 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, and therefore we know that the corresponding U-matrix structures are different. 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 $\{\langle f, g\rangle : g \in B_{\mathrm{f}}(1/2)\}$, where f is defined as

$$f(x) = (1-\cos(8\pi x))\cdot\mathbb{1}_{[0,1]}(x), \qquad x \in [-1, 1]. \tag{5.1}$$

The function f is reconstructed as follows: Let $U := [(B_{\mathrm{f}}(2^{-1}), \rho), (B_2, \tau)]$ 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 standard ordering. Next let $\Omega \subset \mathbb{N}$ denote the set of subsamples from $B_{\mathrm{f}}(2^{-1})$ (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

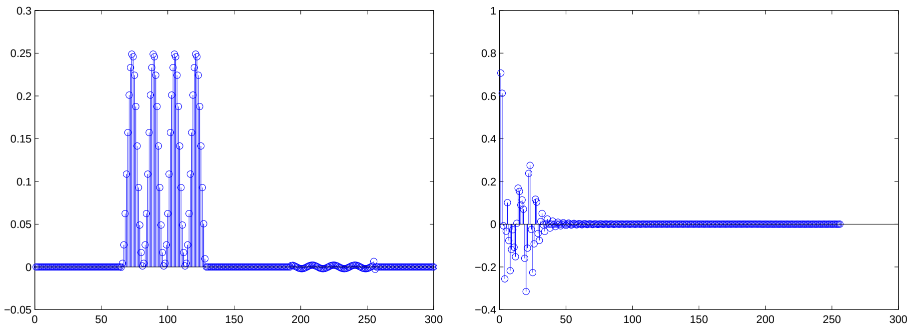
$$\min_{x \in \ell^1(\mathbb{N})} \|x\|_1 \quad \text{subject to} \quad P_\Omega U x = P_\Omega \hat{f}. \tag{5.2}$$

Since the optimisation problem is infinite dimensional we cannot solve it numerically so instead we proceed as in [2] 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

$$\min_{x \in \mathbb{C}^R} \|x\|_1 \quad \text{subject to} \quad P_\Omega U P_R x = P_\Omega \hat{f}. \tag{5.3}$$

We shall be using the SPGL1 package [5] to solve (5.3) numerically. We focus on two choices of reconstruction bases:

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



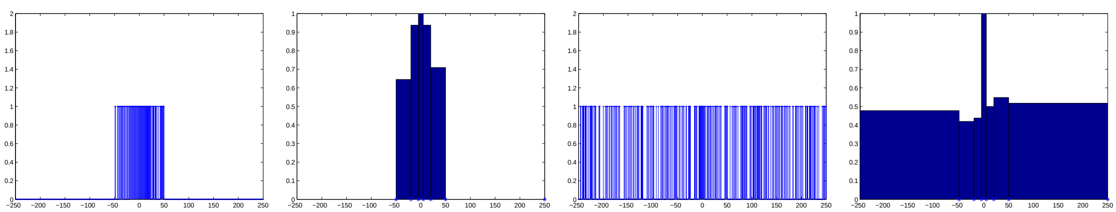
(a) First 300 coefficients (using a leveled ordering) of f in a Daubechies4 boundary wavelet expansion with J=6.

(b) First 100 coefficients (using a natural ordering) of f in a Legendre polynomial expansion.

1. $B_2 = B_{\mathrm{bw}}$ with Daubechies4 boundary wavelets, τ is a leveled ordering.
2. $B_2 = B_{\mathrm{p}}$ 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.

Figure 4: Two sampling patterns and their corresponding histograms.



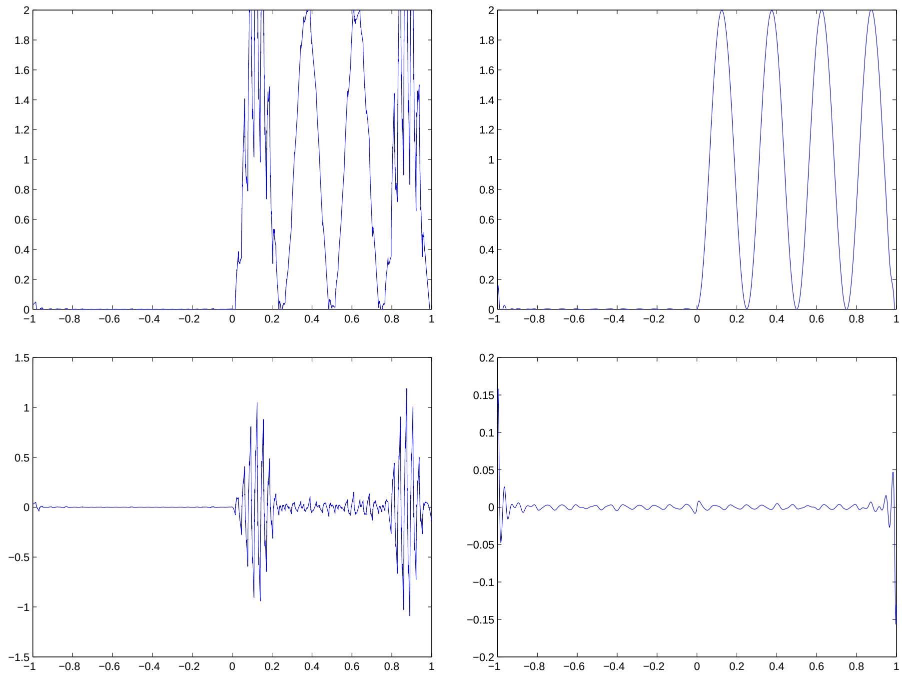
(a) Sampling Pattern A (b) Histogram for Pattern A (c) Sampling Pattern B (d) Histogram for Pattern B

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

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

If we were to sample all the 501 coefficients then we would achieve a highly accurate reconstruction from both bases[13]. 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. 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.

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



(a) Wavelet Reconstruction (b) Polynomial Reconstruction

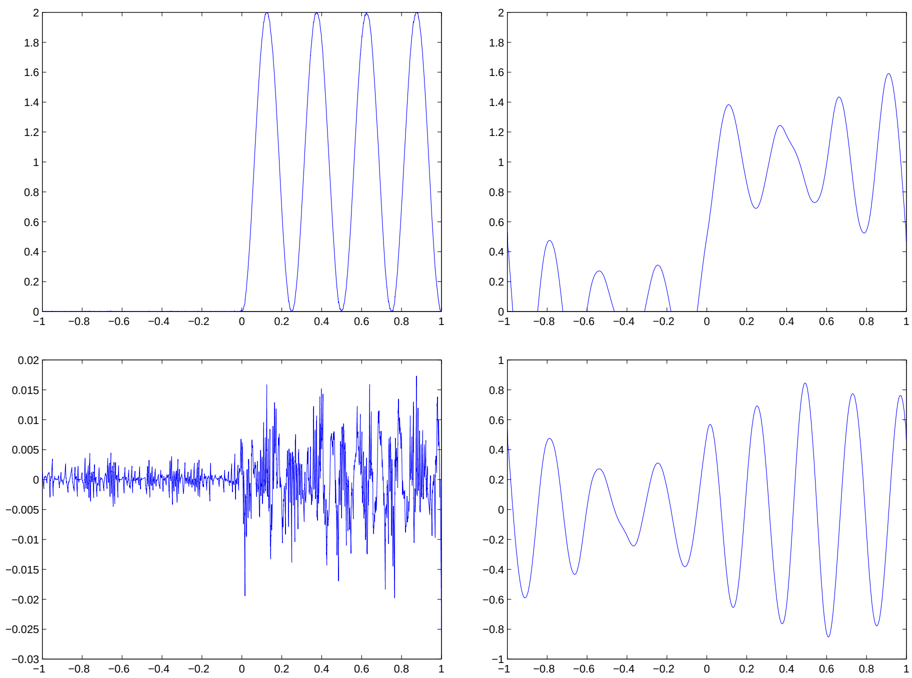
Let us first consider what happens when we use subsampling pattern A, which is shown in Figure 5. We first look at the wavelet reconstruction, which has an L^1 error of 1.52×10^{-1}. The reconstruction fails to reconstruct the smoothness of f, with the first and fourth peaks being particularly jagged. 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 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

[13] For all our reconstructions we will be using $R = 1024$.

Figure 6: Reconstructions from Pattern B with errors.



(a) Wavelet Reconstruction

(b) Polynomial Reconstruction

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.

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 and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Technical report NA2011/02, DAMTP, University of Cambridge*, 2011.

[3] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Preprint*, 2013.

[4] 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.*, (to appear).

[5] E. Berg and M. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal of Scientific Computing*, 31(2):890–912, 2008.

[6] E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[7] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.

[8] 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.

[9] 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.

[10] A. Cohen, I. Daubechies, and P. Vial. Wavelet bases on the interval and fast algorithms. *Journal of Applied and Computational Harmonic Analysis*, 1993.

[11] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41(7):909–996, 1988.

[12] 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.

[13] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[14] Y. C. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[15] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[16] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhuser, 2013.

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

[18] 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.

[19] F. Krahmer and R. Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Process.*, 23 (2):612–622, 2014.

[20] L. J. Landau. Bessel functions: monotonicity and bounds. *J. London Math. Soc. (2)*, 61(1):197–215, 2000.

[21] 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.

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

[23] 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.

[24] S. Mallat. *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition, 2009.

[25] 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.

[26] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, Cambridge, England, 1944.