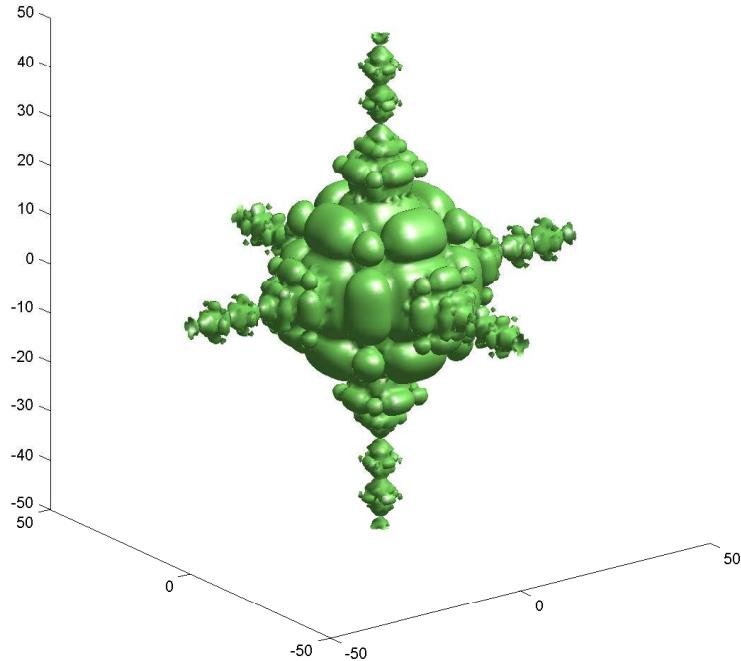
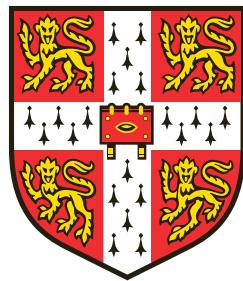


Coherence Structures in Continuous Compressed Sensing with Applications to Helium Atom Scattering



Alexander Daniel Jones

Robinson College
University of Cambridge



This thesis is submitted for the degree of
Doctor of Philosophy
September 2015

Abstract

Coherence, along with sparsity and uniform random sampling, have long been important concepts in discrete Compressed Sensing (CS) since its introduction. In recent years, work by Adcock, Hansen, Poon and Roman has extended this theory to the **continuous case of infinite-dimensional change of basis matrices**. However, in doing so, the concepts underpinning the original CS theory were readdressed. This lead to the notion of coherence being refined to that of local coherence. This thesis begins by giving an overview of some of the current notions of local coherence presented in the field leading up to a focus on asymptotic incoherence, which has been identified as a necessary tool in Continuous CS.

The thesis next presents a new general theory for asymptotic incoherence. Asymptotic decay rates are proven to be of the order N^{-1} for all separable Daubechies wavelet - Fourier pairs in any dimension, **where N denotes the numbers of leading rows/columns omitted**. It is shown that this attains the theoretical limit for any orthonormal basis pair as a power of N . These results allow for established $s \log N$ sampling rates from traditional CS to be brought forward to the continuous case, **where s denotes the sparsity of the signal being recovered**. Optimal asymptotic decay rates are also established for Legendre polynomial - Fourier basis pairs.

The latter part of the thesis documents how the benefits of continuous CS have been brought to the field of Helium Atom Scattering (HAS) in a collaboration between the author and the Surfaces, Microstructure and Fracture research group in Cambridge. Using the spin echo technique developed by the group, continuous CS has been used to reconstruct surface phonon spectra from ${}^3\text{He}$ polarisation data for the first time. Furthermore, the continuous nature of the approach has allowed for the reconstruction of data describing the molecular processes on the surface of materials. This work not only saves time collecting data from the scattering apparatus, but also pushes the limits of what can be measured from existing materials.

Statement of Originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

Chapter 1 consists of a review of the relevant literature, apart from Section 1.7 covering the main results on coherence structures.

Chapters 2 and 3 is based on work on the papers [1] and [2] respectively. Except where explicitly stated, it is original work. The same applies to the conclusions in Chapter 5.

Chapter 4 is based on the paper [3] and is collaborative work with Anton Tamtögl, Irene Calvo-Almazán (from the Surfaces, Microstructure and Fracture (SMF) group at the Cavendish site) and Anders Hansen with myself as main author. My contribution to this work was implementing compressed sensing methods for reconstructing phonon spectra and diffusion data and developing the application of continuous compressed sensing to the full loop shown in Figures 4.14 and 4.1. Sampling data for the gold and cobalt phthalocyanine phonon spectra and data for the synthetic 2D wavelength intensity function in Figure 4.5 was provided by Anton Tamtögl.

Acknowledgements

Firstly I would like to thank my supervisor Anders Hansen for all his support, both technical and moral, throughout my time as a PhD student. I could not have asked for a better supervisor. Even during times of great confusion, the way forward has been made clear meeting after meeting.

Next I would like to thank Ben Adcock for his assistance with writing papers and for some very useful questions. I would also like to express my appreciation for all the people in the AFHA group at DAMPT, in particular Clarice Poon, Milana Gataric and Alex Bastounis for their help and support. In addition I would like to express my gratitude to Daniel Holland for introducing me to compressed sensing and for guiding me in the early days of my PhD.

I have to say a big thank you to all those at the Surfaces, Microstructure and Fracture group at the Cavendish site in Cambridge, who have given me a very good welcome during this past year. In particular I would like to thank Anton Tamtögl, Irene Calvo-Almazán and Bill Allison for their time and effort throughout. They have been a pleasure to work with and I hope that collaboration between the two departments will continue well into the future.

Next I shall like to express my gratitude to those outside my academic work. Firstly, I would like to thank all of colleagues at the CCA, in particular Eoin Devane and Luca Calatroni whom I have shared an office with for the best part of 4 years.

My friends at Robinson College also deserve a mention. A big shout out to all those at RCVGS who have brought me joy for the past 5 years; in particular Richard Allitt, Dylan Ede, Ryutaro Ikeda, Jian Sian Poh, and Nicholas Tomlinson.

I would also like to thank Hu Xin for providing a very different perspective on PhD life from the other side of the world.

Lastly I would like to thank my parents Gary and Janet for their endless care and support.

Table Of Contents

Notation & Conventions	viii
1 Background and Overview	1
1.1 A Beginner's Guide to Compressed Sensing	1
1.1.1 Sampling and Coherence	2
1.2 Applications and Extensions	4
1.2.1 DFT and Problems of Fourier Type	4
1.2.2 Reconstruction Bases	7
1.3 Two Ways to Read This Thesis	8
1.4 Continuous CS	9
1.4.1 The Coherence Barrier	10
1.4.2 Overcoming the Coherence Barrier	11
1.4.3 Relation to the Local Coherences	13
1.5 Basic Framework for Local Coherence Analysis	14
1.6 Some Related Results	16
1.6.1 Discrete Coherence Results	16
1.6.2 Sparsity's Role	18
1.7 Overview of the Thesis and its Contributions	20
1.7.1 One Dimensional Coherence Bounds	21
1.7.2 Multidimensional Coherence Bounds	22
1.7.3 Optimality Results	26
1.7.4 Other Results	27
1.7.5 How this Work Developed	27
2 One Dimensional Cases & Optimality Results	29
2.1 Bases and Orderings	29
2.1.1 Fourier Basis	29

2.1.2	Legendre Bases	30
2.1.3	Standard Wavelets	30
2.1.4	Periodic Wavelets	31
2.1.5	Boundary Wavelets	32
2.1.6	A Few Comments on the Wavelet Bases	33
2.2	1D Fourier-Wavelet Case	33
2.3	1D Fourier-Polynomial Case	39
2.4	The Impact of Incoherence on Sampling Strategies	48
2.5	Proving Optimality of the Coherence Bounds	51
2.6	Theoretical Limits	54
2.7	Alternative Notions of Optimality	57
2.8	Using Asymptotic Incoherence to Derive Standard Sampling Rates	60
3	Multidimensional Cases	65
3.1	Tools for Finding Optimal Orderings	65
3.2	Multidimensional Tensor Cases	67
3.2.1	General Estimates	67
3.2.2	Fourier-Wavelet Case	74
3.2.3	Fourier-Polynomial Case	74
3.2.4	Examples of Hyperbolic Orderings	74
3.2.5	Plotting Tensor Coherences	79
3.3	Multidimensional Fourier - Separable Wavelet Case	79
3.3.1	Defining the Separable Wavelet Basis	79
3.3.2	The Coherence Bounds	81
3.3.3	Ordering the Fourier Basis	83
3.3.4	Examples of Linear Orderings - Linear Scaling Schemes	88
3.3.5	2D Separable Incoherence Plots	89
3.3.6	Semi-Hyperbolic Orderings: Finishing off the Separable Case	89
3.3.7	Optimal Orderings and Wavelet Smoothness	95
3.3.8	Hierarchy of Semihyperbolic Orderings	96
3.3.9	3D Separable Incoherences	97
3.3.10	Comments on Deriving the slogN Rate in Multiple Dimensions	99
3.4	Asymptotic Incoherence and Two Dimensional Compressed Sensing	100
3.4.1	Demonstrating the Benefits of Multilevel Subsampling	102
3.4.2	Tensor vs Separable - Finding a Fair Comparison	102

4 Compressed Sensing of Helium Atom Scattering Spectra	107
4.1 Structure of this Chapter	108
4.2 Background of Helium Spin-Echo Part 1 - Measurements and Phonons .	109
4.2.1 Bravais Crystals, Laue Equations and the Reciprocal Lattice . .	111
4.2.2 Phonons and Changes in Helium Energy	113
4.2.3 Solenoid Currents and Spin Phase	115
4.2.4 The Fourier Slice Theorem	117
4.3 Applying CS to Phonon Detection	119
4.3.1 How this Work Developed	119
4.3.2 Discrete CS Approach	119
4.3.3 Quick Description of Multilevel Sampling	121
4.3.4 Continuous CS for Phonon Spectra	122
4.3.5 Wavelet Bases	123
4.3.6 Simulated 1D Example	126
4.3.7 Real Phonon Spectrum	127
4.3.8 Comparing CS Techniques	129
4.4 Background of Helium Spin-Echo Part 2 - Molecular Diffusion	131
4.4.1 Scattering Cross Sections and the Van Hove Formalism	131
4.4.2 Extension to HeSE	133
4.4.3 The Full Loop - From Wavelength to Energy	134
4.4.4 Why One Should Use Continuous CS for Diffusion	138
5 Conclusions & Future Work	141
5.1 Developments with Compressed Sensing	141
5.1.1 Tools for Identifying Good Reconstruction Bases	141
5.1.2 Directions Forward for Local Coherence Analysis	142
5.2 Developments with Helium Atom Scattering	143
A Hyperbolic Orderings in the Separable Case	145

Notation & Conventions

The set \mathbb{N} denotes the set of natural numbers *excluding zero*.

For a finite set S , the expression $|S| \in \mathbb{N} \cup \{0\}$ refers to the number of elements in S .

For a Banach space $(W, \|\cdot\|)$, the space $\mathcal{B}(W)$ denotes the space of all bounded linear maps from $(W, \|\cdot\|)$ to itself.

Asymptotic Notation

For $f, g : \mathbb{N} \rightarrow \mathbb{R}_{>0}$:

$$\begin{aligned} f(N) = \mathcal{O}(g(N)) &\Leftrightarrow \exists C > 0 \text{ s.t. } f(N) \leq C \cdot g(N) \quad \forall N \in \mathbb{N}, \\ f(N) = o(g(N)) &\Leftrightarrow \forall C > 0 \ \exists N' \in \mathbb{N} \text{ s.t. } f(N) \leq C \cdot g(N) \quad \forall N \geq N', \\ f(N) = \Theta(g(N)) &\Leftrightarrow \exists C_1, C_2 > 0 \text{ s.t. } C_1 \cdot g(N) \leq f(N) \leq C_2 \cdot g(N) \quad \forall N \in \mathbb{N}. \end{aligned} \tag{1}$$

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

$$f \lesssim g \Leftrightarrow \exists C > 0 \text{ s.t. } f(s) \leq C \cdot g(s) \quad \forall s \in S. \tag{2}$$

If the set S is not explicitly stated then (2) is meant to hold for all variables explicitly written in the equation where \lesssim is used (where the range of each variable should be clear from the context).

Chapter 1

Background and Overview

1.1 A Beginner’s Guide to Compressed Sensing

Suppose we have the simple matrix-vector equation

$$y = Ax, \quad x, y \in \mathbb{C}^N, \quad A \in \mathbb{C}^N \times \mathbb{C}^N, \quad (1.1)$$

where the matrix A represents a linear transform that takes the vector (or data), represented by x , into some (sampling) data y that can be detected. The term *sensing* refers to the process of taking samples of y , using the *sensing matrix* A , which is only way of extracting information from x , with the goal of obtaining full knowledge of x . Assuming the matrix A is invertible, full knowledge of y will allow the reconstruction of x .

Compressed sensing (CS), introduced by Candès, Romberg & Tao [4] and Donoho [5], has been one of the major achievements in applied mathematics in the last decade [6, 7, 8, 9, 10]. The goal of CS is to obtain a good reconstruction of x from *partial* knowledge of y . In this case the term “partial knowledge” refers to knowledge of a fraction of the coefficients $\{y_i : i \in \Omega\}$ where $\Omega \subset \{i = 1, \dots, N\}$. This would mean that the problem (1.1) would become

$$P_\Omega y = P_\Omega Ax, \quad x, y \in \mathbb{C}^N, \quad A \in \mathbb{C}^N \times \mathbb{C}^N, \quad (1.2)$$

where P_Ω denotes the projection onto the indices in Ω . Clearly we cannot use the partial information alone to reconstruct the full vector x ($P_\Omega A$ is not invertible) and additional information must be used to make the problem feasible.

In CS the extra information (that one assumes) comes from x being known to be

sparse. In it's most simple form this is defined as

$$s(x) := |\{x_j \neq 0 : j = 1, \dots, N\}|. \quad (1.3)$$

This assumption can be imposed by adding a minimising objective to (1.2), leading to the problem of finding the (hopefully unique) solution to

$$\min_{x' \in \mathbb{C}^N} s(x'), \quad \text{s.t. } P_\Omega y = P_\Omega A x'. \quad (1.4)$$

In practice this problem is numerically intractable [11] therefore in practice the sparsity term is replaced with an ℓ^1 term:

$$\min_{x' \in \mathbb{C}^N} \|x'\|_1 = \sum_{i=1}^N |x'_i|, \quad \text{s.t. } P_\Omega y = P_\Omega A x'. \quad (1.5)$$

This problem can then be solved using various convex optimisation packages (in this thesis we rely on the SPGL1 package [12, 13] which targets this particular class of problem).

1.1.1 Sampling and Coherence

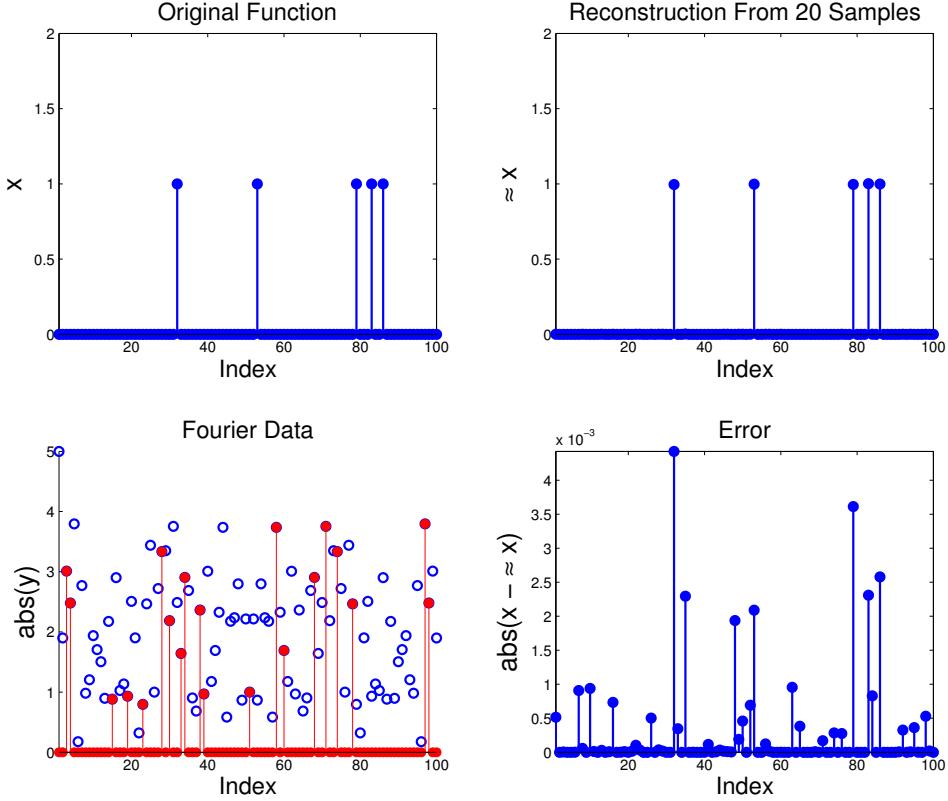
Next comes the question; how should one choose the subset of samples $P_\Omega y$ so that we can get the best reconstruction of x , given that we only know $s(x)$? Since the positions of the non-zeros entries of x are unknown, this strongly suggests that *uniformly random sampling* is the answer.

Next we have the question: when does this strategy work? This can be broken down into two sub-questions: for what kind of sampling matrices A does this approach work best for and how many samples do we have to take?

The answer to the first question can be deduced by looking at particular cases. If A was concentrated along the diagonal, e.g. $A = I$, then reconstructing x means sampling every y_j such that $x_j \neq 0$. Since the j 's are chosen uniformly at random the probability of this happening is negligible and the reconstruction would be poor. This suggests that the matrix A should be spread out rather than concentrated, but how to make this all precise?

To keep things simple, we shall assume that A represents a change of basis matrix from the standard coordinate basis to another orthonormal basis (this is the same as

Figure 1.1: Simple discrete CS example using an $A = DFT$ sensing matrix at a resolution of $N = 100$ with sparsities ($s(x) = 5$) chosen at random. 20 uniform Samples (indicated in red) are taken from the Fourier data to reconstruct the original data.



saying A is an *isometry*). Because A is an isometry, its columns are normalised and we can use the absolute size $|\cdot|$ to measure how flat the matrix A is:

Definition 1.1.1. *The quantity*

$$\mu(A) = \sup_{i,j \in \{1, \dots, N\}} |A_{ij}|^2, \quad A \in \mathbb{C}^{N \times N}, \quad N \in \mathbb{N}, \quad (1.6)$$

is called the *Coherence of A* . For an isometry A the coherence $\mu(A)$ takes values between N^{-1} (perfectly incoherent [15]) and 1 (perfectly coherent).

Perfect incoherence is the case where A is uniformly flat and $|A_{i,j}| = N^{-1}$ for every i, j .

Next we look at the number of samples. The classical idea of CS is that Ω should be chosen uniformly at random. In this case the number of samples $m = |\Omega|$ required

to reconstruct x with probability $1 - \epsilon$ should satisfy [14, 15]

$$m \gtrsim N \cdot \mu(A) \cdot s(x) \cdot \log(\epsilon^{-1}) \cdot \log(N), \quad N \in \mathbb{N}. \quad (1.7)$$

Therefore, the smaller the sparsity $s(x)$ and coherence $\mu(A)$, the fewer the number of samples of y that is required to reconstruct x successfully.

Figure 1.1 shows a stereotypical example of where this approach works best. Here we have $s(x)$, where the sparsities are chosen uniformly at random and 20 samples are taken from y uniformly at random.

1.2 Applications and Extensions

1.2.1 DFT and Problems of Fourier Type

Our first example is the Discrete Fourier Transform (DFT) matrix

$$A_{jk} = (\text{DFT})_{jk} = \frac{1}{\sqrt{N}} \exp\left(2\pi i \frac{(j-1)(k-1)}{N}\right), \quad j, k = 1, \dots, N, \quad N \in \mathbb{N}. \quad (1.8)$$

The DFT matrix has perfect incoherence $\mu(A) = N^{-1}$, making it an ideal target for CS. Moreover it has relevance to many physical applications:

For spectroscopy problems of Fourier type, (e.g. Nuclear Magnetic Resonance (NMR) [16, 17], Neutron / Helium spin echo [18, 19], X-ray Diffraction [20]) this matrix arises from a continuous Fourier transform in the sensing model, namely to reconstruct $f \in L^2(\mathbb{R}^d)$, $d \in \mathbb{N}$ from samples of its continuous Fourier Transform (FT):

$$g(\omega) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \omega \cdot x} dx, \quad \omega \in \mathbb{R}^d. \quad (1.9)$$

The function f could be obtained by the inverse Fourier transform of g . For a general function f this would require knowing $g(\omega)$ at every point $\omega \in \mathbb{R}^d$ which is unrealistic.

In many examples in practice however, the function f is treated as a periodic (or compactly supported) function over some fixed region, say $[a, b]$ for the one dimensional case. This is convenient because it permits changing the problem to one of handling a Fourier transform to that of handling a Fourier series expansion:

$$\begin{aligned}
f(x) &= \sum_{k \in \mathbb{Z}} \langle f, \exp_{k,\epsilon} \rangle \exp_{k,\epsilon}(x), \quad x \in [a, b], \\
\langle f, \exp_{k,\epsilon} \rangle &= \int_{\mathbb{R}} f(x) \overline{\exp_{k,\epsilon}(x)} dx = g(k\epsilon) \epsilon^{1/2}. \tag{1.10} \\
\exp_{k,\epsilon}(x) &= \epsilon^{1/2} \exp(2\pi i \epsilon k \cdot x), \quad k \in \mathbb{Z}, \quad \epsilon = (b-a)^{-1},
\end{aligned}$$

The upshot of (1.10) is that we now only need the values $g(k\epsilon)$, $k \in \mathbb{Z}$, to obtain the function f , rather than $g(\omega)$, $\omega \in \mathbb{R}$.

Typically the next step is to truncate the Fourier series expansion, meaning that one makes the approximation

$$\begin{aligned}
f(x) &\approx \tilde{f}_N(x) = \sum_{k=-N}^N \langle f, \exp_{k,\epsilon} \rangle \exp_{k,\epsilon}(x), \quad x \in [a, b], \\
|f(x) - \tilde{f}_N(x)|^2 &= \sum_{|k|>N} |\langle f, \exp_{k,\epsilon} \rangle|^2, \tag{1.11}
\end{aligned}$$

for some fixed $N \in \mathbb{N}$. The problem is now feasible as only finitely many data points k are required to determine \tilde{f}_N . The DFT matrix now arises from evaluating (1.11) on a uniformly spaced grid of points $x_{j,N} = a + \frac{b-a}{2N+1} j$, $j = 0, \dots, 2N$ in $[a, b]$.

Figure 1.2 shows an application of CS using a DFT sensing matrix, however this time it is used to construct a smooth continuous function. Here we have a synthetic spectrum made up of Lorentzian functions (often used to model NMR spectra [16]):

$$L(x)_{\lambda,\omega_0} = \frac{\lambda}{\lambda^2 + (x - \omega_0)^2}, \quad x \in \mathbb{R}, \tag{1.12}$$

and work with a resolution of around 500 points. While we can subsample to some degree with this example, once the sampling rate drops below 50% the peaks begin to collapse.

There are now a few issues that were not present in the example of Figure 1.1. Firstly is the notion of sparsity here: the function is technically non-zero everywhere, although it is close to zero over the majority of the domain. This means that we could try to define a modified version of sparsity based on a threshold:

$$s_\epsilon(x) := |\{j = 1, \dots, N : |x_j| > \epsilon\}|. \tag{1.13}$$

The problem with this approach (and with 1.3) is that these notions of sparsity does not

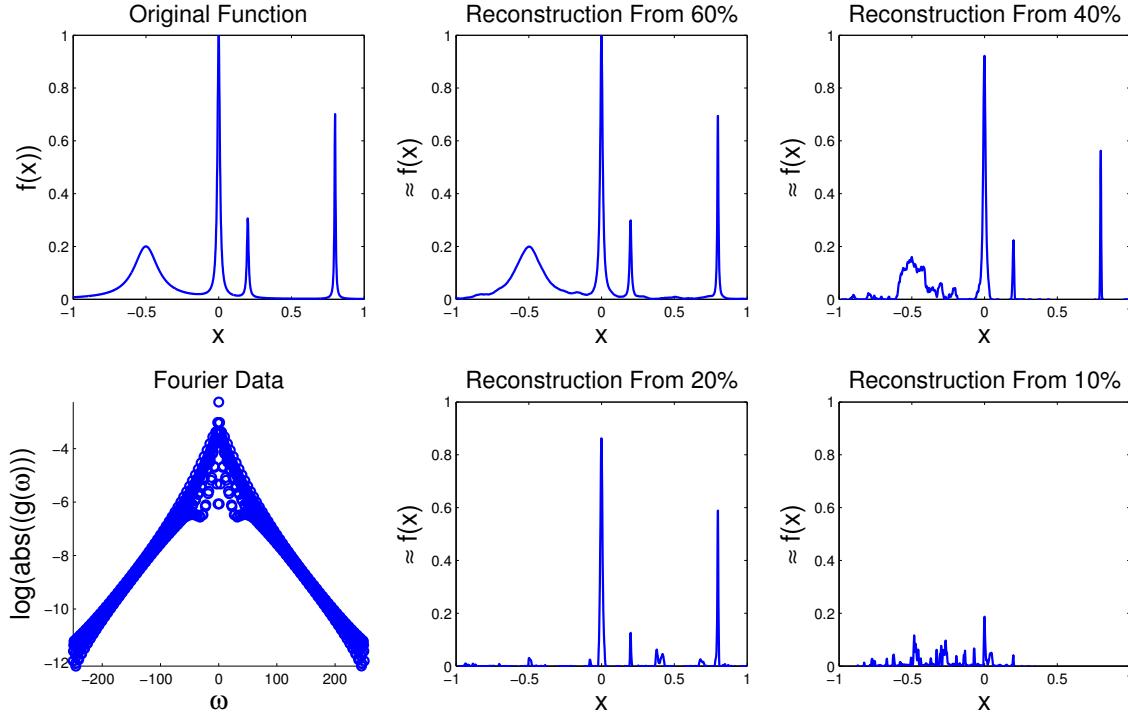


Figure 1.2: Traditional CS (with uniform random sampling) applied to a synthetic spectra. Samples are taken from the frequency range $[-250, 250]$ at 0.5 point spacings. Observe that the wider peak on the left breaks down very quickly under subsampling due to lack of sparsity.

scale well with resolution. If one doubles the resolution, the resulting sparsity will also double therefore, according to (1.7), even in the case of perfect incoherence, the number of samples required to reconstruct the function will also double, which is a concern since heuristically the function has not gotten more complex.

Therefore, despite the DFT matrix having perfect coherence, it has the following drawbacks:

1. There is no way of handling cases with lack of sparsity **since there is no sparsifying basis or transform.**
2. Since the DFT is a square matrix, the resolution of the reconstruction is fixed by the **total number** of frequencies **that are being subsampled from**. This implicitly impacts the sparsity since it becomes dependent on frequency.
3. By discretising (1.11) we have potentially lost information.
4. **The reconstruction is sampled on a uniform grid which may become distorted if the reconstruction needs to be transformed nonlinearly** (as in the work on Helium

Spin Echo in Chapter 4).

1.2.2 Reconstruction Bases

To get around the issue of lack of sparsity, one can try to transform the target data x so that it becomes sparse. Formally, this means introducing another matrix $W \in \mathbb{C}^N \times \mathbb{C}^N$, called the *sparsifying transformation*, that is expected to make the transformed data Wx sparse. We shall assume that, like the sensing matrix A , the sparsifying transformation is an isometry, namely it can be represented by an orthonormal basis. Examples of sparsifying bases include:

- [Wavelets] This huge class of bases all have a notion of scaling (resolution) levels. Formally, a wavelet basis has a scaling function ϕ and mother wavelet ψ that generate the entire basis. Examples of wavelet bases include Haar wavelets (arguably the simplest of all wavelet bases), Daubechies wavelets (minimum support given number of vanishing moments of the mother wavelet), Morlets (Windowed frequency analysis, infinite support), etc. For more details see [21, 22] or Section 2.1.3.
- [Polynomials] In the same spirit as Taylor series. Useful for smooth functions with rapidly decaying Taylor expansions. In this thesis a Legendre polynomial basis is considered (see Section 2.1.2).

There are also other families that are not orthonormal but nonetheless can provide sparse representations. These include biorthogonal wavelets (e.g. CDF 5/3 wavelet [23] used in JPEG 2000), frames (e.g. shearlets [24]) and total variation (TV) methods [25], but are beyond the scope of this thesis are a potential area of future research.

The benefit of using a sparsifying transformation is that it can depend on the type of data we are trying reconstruct. Want to reconstruct smooth data in one dimension? Use a high order Daubechies wavelet. Is the data mostly flat with a few discontinuities? Try Haar or TV.

Adding a sparsifying transform to (1.1) gives us (for $z = Wx$)

$$y = AW^{-1}z, \quad z, y \in \mathbb{C}^N, \quad A, W \in \mathbb{C}^N \times \mathbb{C}^N. \quad (1.14)$$

This is actually the same as (1.1) except the matrix A is replaced by AW^{-1} . Therefore we can solve this problem in the same manner:

$$\min_{z' \in \mathbb{C}^N} \|z'\|, \quad \text{s.t. } P_\Omega y = P_\Omega A W^{-1} z', \quad (1.15)$$

where the minimiser z^* can then be used to derive $x = W^{-1}z$. Rephrasing this problem in terms of x is the same as solving

$$\min_{x' \in \mathbb{C}^N} \|Wx'\|, \quad \text{s.t. } P_\Omega y = P_\Omega A x'. \quad (1.16)$$

Likewise the same sampling guarantees are retained: the number of uniform samples required to obtain an exact reconstruction of x with probability greater than $1 - \epsilon$ must satisfy

$$m \gtrsim N \cdot \mu(AW^{-1}) \cdot s(x) \cdot \log(\epsilon^{-1}) \cdot \log(N), \quad N \in \mathbb{N}. \quad (1.17)$$

At this point a new problem arises, what is the size of $\mu(AW^{-1})$? One example to consider is A being a DFT matrix and W being a Haar wavelet transform. If one assumes for simplicity that $N = 2^j, j \in \mathbb{N}$ and we have $j - 1$ wavelet levels (i.e. a full wavelet decomposition) then one can show¹ that $\mu(AW^{-1}) = 1$, i.e. perfect coherence which prevents any kind of useful compression guarantee using (1.17).

However, one does see that in practice one can compress this problem using *non-uniform sampling* and notions of *local coherence*, something we shall come onto in the next few sections. Even so, assuming these issues can be addressed, not all of the problems we listed in the previous section have been resolved by using reconstruction bases. In particular the reconstruction is still inherently discrete and is locked to the frequency range we are sampling from. While this is less of a problem for many discrete models, for most physical problems of Fourier type that have a continuous model a continuous reconstruction is highly desirable.

1.3 Two Ways to Read This Thesis

At this point there are two ways to proceed with this thesis. Those interested in the general theory of continuous CS and coherence structures should continue reading the thesis in a linear fashion. In this case Section 1.7 gives an overview of the results on coherence structures.

For those who would prefer to go straight to the application of continuous CS to

¹This is because both the discrete Fourier basis and the Haar wavelet basis share the constant vector $N^{-1/2}(1, \dots, 1)$.

Helium Atom Scattering (HAS) we recommend giving the continuous CS section below a quick skim through before moving onto Chapter 4 which focuses on using continuous CS as a tool whilst avoiding the technical details where possible.

Chapter 5 discusses the conclusions of both the coherence results and the applications to HAS.

1.4 Continuous CS

Let us return to the Fourier series expansion in (1.10). In that section we first discretised the problem and then applied a sparsifying matrix transform that represented a reconstruction basis. However, in the same manner of the Fourier expansion, one can try approximating f in terms of a new *continuous* reconstruction basis σ_n , $n \in \mathbb{N}$, without having to discretise the problem at all:

$$f(\lambda) = \sum_{n=1}^{\infty} \langle f, \sigma_n \rangle \sigma_n(\lambda), \quad \lambda \in [a, b]. \quad (1.18)$$

Apart from the benefit of the keeping the problem continuous, one also has the freedom to *choose* which basis σ_n to work with, making the approach more versatile than a straight DFT approach. Continuous reconstruction bases are as simple if not easier to construct as discrete ones. In fact all of the reconstruction bases mentioned in Section 1.2.2 have either continuous origins or, at the very least, a continuous analogue. For example a Daubechies wavelet is constructed through its continuous (trigonometric) polynomial filter [26].

Since we are still sampling data that corresponds to Fourier coefficients of f , it is impossible to exclusively work with the choice of basis σ_n . Instead we convert Fourier series coefficients into coefficients in the basis σ_n . This is achieved by working with the change of basis matrix between the two:

$$B_{k,n} = \langle \sigma_n, \exp_{k,\epsilon} \rangle, \quad n \in \mathbb{N}, \quad k \in \mathbb{Z}. \quad (1.19)$$

Using this matrix to reconstruct the coefficients $\langle f, \sigma_n \rangle$, $n \in \mathbb{N}$ we can then use (1.18) to approximate $f(\lambda)$.

1.4.1 The Coherence Barrier

It was shown in [27] that the result of (1.7) could be extended to cover this case, however a new problem presented itself; does the coherence of the matrix B in (1.19) make sense in this context?

Of course we can still define coherence in this case as follows ($U \in \mathcal{B}(\ell^2(\mathbb{N}))$):

$$\mu(U) = \max_{i,j \in \mathbb{N}} |U_{ij}|^2, \quad (1.20)$$

however there is now a fundamental issue with this definition. In the finite-dimensional case the coherence of an isometry W was a measure of how flat W was; what does it mean for an infinite dimensional matrix to be flat?

Suppose that the matrix $U : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ is an isometry. By the definition of an isometry we see that the first column of U is a unit vector in $\ell^2(\mathbb{N})$. If U was perfectly flat then every entry of the first column would have the same size, implying that the column has either infinite or zero norm, contradicting it being a unit vector.

Furthermore, because of this issue, any reconstruction basis will lack incoherence and consequently the result of (1.7) becomes harder to apply in general.

In the search for our alternative, examples often guide the way. An important example of this is using a periodised wavelet basis (see Section 2.1.4) with an initial resolution level spanning the entire reconstruction interval. In this case one can show that $\mu(U) = 1$ which is the worst possible scenario. This phenomena has been coined the *Coherence Barrier* [15]. This is the same problem that we encountered in the discrete case in Section 1.2.2.

As Figure 1.3 reveals, although such measurement matrices are clearly not flat, it does have some kind of structure. For example, the high coherence terms (large matrix entries) are concentrated around a submatrix of U .

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

$$\pi_N(x)_i := \begin{cases} 0 & i \neq N \\ x_i & i = N \end{cases}, \quad R_N(x)_i := \begin{cases} 0 & i < N \\ x_i & i \geq N \end{cases}, \quad (1.21)$$

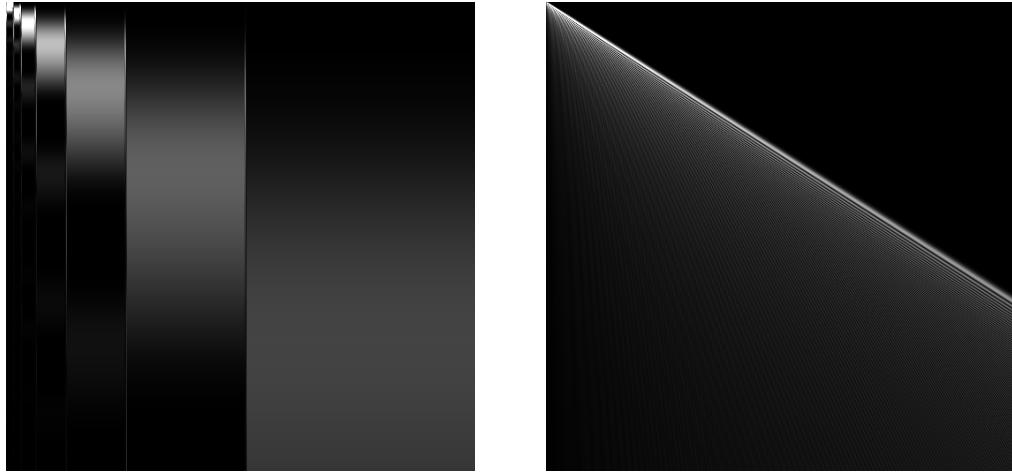


Figure 1.3: Plots of the absolute values of the entries of the compression matrix U corresponding to Fourier sampling with Daubechies6 boundary wavelets (left) and Legendre polynomials (right). Light regions correspond to large values and dark regions to small values. $U_{1,1}$ corresponds to the top-left entry. Notice that if these matrices were incoherent, then these figures would have been monotone. This means an alternative to coherence is needed to study infinite dimensional compression matrices.

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

$$\mu(\pi_N U), \mu(U\pi_N), \mu(R_N U), \mu(UR_N). \quad (1.22)$$

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

Now that we have established how one can treat coherence in an infinite dimensional setting, we next consider how this influences CS theory.

1.4.2 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 at different rates over different frequency ranges. This was recently confirmed by mathematical analysis in [15, 28, 29]. The key to their work was to work with the three concepts of *sparsity in levels*, *multi-level sampling* and *local coherence* – and prove recovery estimates akin to 1.7 under these more general

settings.

Definition 1.4.1 (Sparsity in levels). *Let x be an element of either \mathbb{C}^N or $\ell^2(\mathbb{N})$. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \dots < M_r$ and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \dots, r$, where $M_0 = 0$. We say that x is (\mathbf{s}, \mathbf{M}) -sparse if, for each $k = 1, \dots, r$,*

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

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

Definition 1.4.2 (Multi-level sampling scheme). *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r = N$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, and suppose that*

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

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

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

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

Definition 1.4.3. *Let U be an isometry of either \mathbb{C}^N or $\ell^2(\mathbb{N})$. Let $\mathbf{N} = (N_1, \dots, N_{r_1}) \in \mathbb{N}^{r_1}$ and $\mathbf{M} = (M_1, \dots, M_{r_2}) \in \mathbb{N}^{r_2}$ with $1 \leq N_1 < \dots < N_{r_1}$ and $1 \leq M_1 < \dots < M_{r_2}$. In [15] the authors introduced their own definition of local coherence as follows; the $(k, l)^{\text{th}}$ local coherence of U with respect to \mathbf{N} and \mathbf{M} is*

$$\mu_{\mathbf{N}, \mathbf{M}}(k, l) = \sqrt{\mu(P_{N_{k-1}}^{N_k} U P_{M_{l-1}}^{M_l}) \cdot \mu(P_{N_{k-1}}^{N_k} U)}, \quad k = 1, \dots, r_1, \quad l = 1, \dots, r_2, \quad (1.23)$$

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

In [15] a new theory of compressed sensing for changes of bases between infinite-dimensional Hilbert-spaces was introduced based on these assumptions. In this case we solve the following problem: if $x \in \ell^2(\mathbb{N})$ is (s, \mathbf{M}) -sparse and U is an isometry of $\ell^2(\mathbb{N})$

then we hunt for the (hopefully unique) η that solves

$$\min_{\eta \in \ell^2(\mathbb{N})} \|\eta\|_{l^1} \quad \text{subject to} \quad P_\Omega U \eta = P_\Omega U x, \quad (1.24)$$

Linking this up with the previous discussion, the change of basis B in (1.19) can be used as the matrix U in the above.

In practice the matrix $P_\Omega U$ still has an infinite number of columns and has to be truncated on the other side. This means replacing $P_\Omega U$ by $P_\Omega U R_N$ in (1.24) where $N \in \mathbb{N}$ is sufficiently large. For large scale problems (e.g. working with high resolution images) this matrix section is too large to work with directly. Instead, in the Fourier-wavelet cases, there are fast transforms available [30] that rely on multiple Fast Fourier Transforms (FFT) and storing Fourier transform coefficients of the mother wavelet / scaling function using formulae such as (2.7).

In this case, instead of a standard compressed sensing estimate 1.7 determining the total number of measurements, one has the following estimate regarding the local number of measurements m_k in the k^{th} level:

$$m_k \gtrsim (N_k - N_{k-1}) \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r_2} \mu_{\mathbf{N}, \mathbf{M}}(k, l) \cdot s_l \right) \cdot \log(N), \quad k = 1, \dots, r_1. \quad (1.25)$$

When the above condition is satisfied, [15] found that with probability exceeding $1 - \epsilon$, x can be shown to be the unique minimiser to (1.24). In particular, the number of samples m_k needed to be taken in each region $\{N_{k-1} + 1, \dots, N_k\}$ can be inferred through the local sparsities and coherences using the asymptotic relation (1.25).

1.4.3 Relation to the Local Coherences

Now let us look at how the local coherences defined in (1.22) fit into this framework. First we notice that by the definitions of R_N and P_a^b that $\forall a, a', b, b', N \in \mathbb{N}$, $b \geq N$,

$$\mu(P_{N-1}^b U) \leq \mu(R_N U), \quad \mu(U P_{N-1}^b) \leq \mu(U R_N), \quad (1.26)$$

$$\mu(P_a^b U P_{a'}^{b'}) \leq \min(\mu(P_a^b U), \mu(U P_{a'}^{b'})). \quad (1.27)$$

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

$$\mu_{\mathbf{N}, \mathbf{M}}(k, l) \leq \sqrt{\min(\mu(R_{N_{k-1}+1} U), \mu(U R_{M_{l-1}+1})) \cdot \mu(R_{N_{k-1}+1} U)}. \quad (1.28)$$

Therefore, understanding the behaviour of the block coherences $\mu(R_N U), \mu(UR_N)$ enables the bounding of the coherence terms (1.23). Consequently, a viable sampling strategy can be deduced via (1.25). The upshot of working entirely with block coherences is that one can optimise the decay of $\mu(R_N U), \mu(UR_N)$ separately which allows us to decouple the orderings of each basis; something that shall be covered in the next section.

At this point it is worth mentioning whether reducing (1.23) to the weaker bound (1.28) prevents us from giving an optimal bound on the number of samples required. It turns that the using the bound (1.27) does end up in an additional $\log(N)$ factor in (1.25) for the Fourier-wavelet case although this can be overcome using a technique from [31]. How to deal with this complication is covered in Section 2.8.

1.5 Basic Framework for Local Coherence Analysis

At this point it is worth clarifying the basic theoretical framework that Chapters 2 and 3 of the Thesis considers.

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

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

We call (B_1, B_2) a ‘basis pair’.

This setup includes (1.19) for the case $[a, b] = [0, 1]$:

$$\mathcal{H} = L^2([a, b]), \quad B_1 = B_f = \{\chi_{k,\epsilon} : k \in \mathbb{Z}\}, \quad B_2 = \{\sigma_n : n \in \mathbb{N}\}, \quad (1.29)$$

where B_f denotes the one-dimensional Fourier basis defined by

$$B_f(\epsilon) = \{\chi_{k,\epsilon} : k \in \mathbb{Z}\}, \quad \chi_{k,\epsilon}(x) := \sqrt{\epsilon} \exp(2\pi i \epsilon \cdot kx) \cdot \mathbb{1}_{[0,\epsilon^{-1}]}(x). \quad (1.30)$$

We often drop the ϵ in $\chi_{k,\epsilon}$ and just write χ_k . Next we have a deceptively simple but very important definition:

Definition 1.5.1 (Orderings). *Let S be a set. Say that a function $\rho : \mathbb{N} \rightarrow S$ is an ‘ordering’ of S if it is bijective.*

An ordering of a basis typically reflects its intrinsic structure. An example of this

would be ordering the Fourier basis by frequency as follows:

$$(\chi_{0,\epsilon}, \chi_{1,\epsilon}, \chi_{-1,\epsilon}, \chi_{2,\epsilon}, \chi_{-2,\epsilon}, \chi_{3,\epsilon}, \dots), \quad (1.31)$$

This sort of notation gets cumbersome very quickly and only gets worse in higher dimensions. To simplify things we use the following tool:

Definition 1.5.2 (Consistent ordering). *Let $F : S \rightarrow \mathbb{R}$ where S is a set. We say that an ordering $\rho : \mathbb{N} \rightarrow S$ is ‘consistent with respect to F ’ if*

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

Let’s now use this tool to do the work of (1.31):

Definition 1.5.3 (Frequency ordering). *Define $F_f : B_f(\epsilon) \rightarrow \mathbb{N}$ by $F_f(\chi_{k,\epsilon}) = |k|$. Say that an ordering $\rho : \mathbb{N} \rightarrow B_f(\epsilon)$ is a ‘frequency ordering’ if it is consistent with F_f .*

Notice that (1.31) describes a Frequency ordering. Moreover if we swap the positions of $\chi_{k,\epsilon}$ and $\chi_{-k,\epsilon}$ in a frequency ordering we still have a frequency ordering.

We finish this section with a the definition of a change of basis matrix using orderings of the bases.

Definition 1.5.4 (Change of Basis Matrix). *For a basis pair (B_1, B_2) , with corresponding orderings $\rho : \mathbb{N} \rightarrow B_1$ and $\tau : \mathbb{N} \rightarrow B_2$, form a matrix $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ by the equation*

$$U_{m,n} := \langle \tau(n), \rho(m) \rangle. \quad (1.32)$$

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

For the theoretical part of this thesis, the main target of study is the line and block coherences (1.22). We shall use the following simple property of these local coherences throughout Chapters 2 and 3:

Lemma 1.5.5. *Let $U := [(B_1, \rho), (B_2, \tau)]$. Then the line/block coherences*

$$\mu(\pi_N U), \mu(R_N U),$$

are invariant under changes of the ordering τ of the second basis B_2 . Likewise the line/block coherences $\mu(U\pi_N), \mu(UR_N)$ are invariant under changes of the ordering ρ of the first basis B_1 .

Proof. Changing the ordering τ is the same as permuting the columns of U . Since $\mu(\pi_N U), \mu(R_N U)$ by definition take the suprema over all columns of U they are left unchanged. \square

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

Lemma 1.5.6. *Let $U = [(B_1, \rho), (B_2, \tau)]$. Suppose $\mu(\pi_N U) = \Theta(f(N))$ for some decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$. Then $\mu(R_N U) = \Theta(f(N))$. Likewise for $\mu(U\pi_N)$ and $\mu(UR_N)$.*

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

$$\mu(R_N U) = \max_{M \geq N} \mu(\pi_M U) \leq C_2 \max_{M \geq N} f_2(M) = C_2 f(N). \quad (1.33)$$

\square

1.6 Some Related Results

Here we briefly cover some relevant research in the field on local coherence and applications of CS in continuous settings.

Coherence has always held an important place in CS theory and has been given its fair share of attention in academic research. Because of this the coverage here is restricted to the most relevant work on local coherence that the author is aware of.

Besides the original continuous CS work above in [15, 27] that introduced asymptotic incoherence and the related work [32, 31] by the Applied Functional and Harmonic Analysis group at DAMTP, there has been little work on local coherence in the continuous setting between infinite dimensional change of bases.

There is some work [33, 34] on coherence structures of problems involving reconstructing polynomials from pointwise samples but this does not fit into the change of basis framework considered in this thesis and does not use Fourier sampling.

1.6.1 Discrete Coherence Results

We focus on uses of local coherence for changes of bases in a discrete setting with Fourier sampling.

In [35], the authors work with discretised Fourier and Haar wavelet elements φ_k and $h_{n,l}^e$ defined on $\mathbb{N}^N = \mathbb{N}^{2^p}$ as follows:

$$\begin{aligned}\varphi_k(t) &= \frac{1}{\sqrt{N}} e^{2\pi i t k / N}, \quad -N/2 + 1 \leq k \leq N/2 \in \mathbb{Z}, \\ h^0(t) &= 2^{-p/2}, \quad h^1(t) = \begin{cases} 2^{-p/2}, & 1 \leq t \leq 2^{p-1}, \\ -2^{-p/2}, & 2^{p-1} < t \leq 2^p, \end{cases} \\ h_{n,l}^e(t) &= 2^{n/2} h^e(2^n t - 2^p l), \quad 0 < n < p, \quad 0 \leq l < 2^n.\end{aligned}\tag{1.34}$$

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

$$\begin{aligned}|\langle \varphi_k, h_{n,l} \rangle| &\leq \min\left(\frac{6 \cdot 2^{n/2}}{|k|}, 3\pi 2^{-n/2}\right), \\ k &= -N/2 + 1, \dots, N/2 \in \mathbb{Z}, \quad n = 1, \dots, p-1 \in \mathbb{N},\end{aligned}\tag{1.35}$$

Result (1.35) was used to derive the following bound in Corollary 6.4:

$$|\langle \varphi_k, h_{n,l} \rangle| \leq \frac{3\sqrt{2\pi}}{\sqrt{k}}, \quad k = -N/2 + 1, \dots, N/2 \in \mathbb{Z}.\tag{1.36}$$

With the terminology of our paper, the result (1.36) shows that the corresponding discrete change of basis matrix $U_N \in \mathbb{C}^N \times \mathbb{C}^N$ satisfies (ordering the φ_k by frequency c.f Def. 1.5.3)

$$\mu(\pi_N U_N) = \mathcal{O}(N^{-1}).\tag{1.37}$$

They also considered the analogous two-dimensional discrete case using a bivariate (i.e. separable) Haar reconstruction basis: if

$$\varphi_{k_1, k_2}(x, y) = \varphi_{k_1}(x) \cdot \varphi_{k_2}(y), \quad -N/2 + 1 \leq k_1, k_2 \leq N/2,\tag{1.38}$$

denotes the two-dimensional Fourier basis of $\mathbb{C}^N \times \mathbb{C}^N$ and $h_{n,l}^e$ denotes a Haar basis term then (Theorem 6.2 in [35])

$$|\langle \varphi_{k_1, k_2}, h_{n,l}^e \rangle| \leq \min\left(1, \frac{18\pi}{\max(|k_1|, |k_2|)}\right).\tag{1.39}$$

Again in the framework used in this thesis, this bound can be optimised and rephrased

in terms of orderings as follows

$$\mu(\pi_N U) = \mathcal{O}(N^{-1}), \quad (1.40)$$

where the Fourier basis is ordered consistent with the function $F(\varphi_{k_1, k_2}) = \max(|k_1|, |k_2|)$ (namely a Linear Ordering (see Definition 3.3.14)).

The closest results to (1.37), (1.40) in this thesis are for a continuous change of basis matrix between Fourier and (separable) wavelet elements in $L^2([0, 1]^d)$ shown in (1.44) and (1.52)

$$\mu(\pi_N U) = \Theta(N^{-1}), \quad \mu(U\pi_N) = \Theta(N^{-1}). \quad (1.41)$$

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

The methods of proof are also worth comparing. In [35] explicit forms of the discrete Haar basis in Section 2.2 are used to derive (1.36). For higher order Daubechies wavelets such an explicit description is not available and this approach cannot be extended. In this thesis we instead rely on the fact that the entries of U can be viewed as Fourier transforms of the wavelet basis e.g. (2.6). Since Daubechies wavelets were originally constructed through their Fourier transform, this analytical approach is very natural.

1.6.2 Sparsity's Role

A key point regarding the discrete results above is that they are based on the sparsity model and the Restricted Isometry Property (RIP). In particular they only focus on local coherences on the sampling side ($\mu(\pi_N U)$) and leave out the sparsity side ($\mu(U\pi_N)$). This is due to the face that in the work [35] a standard sparsity model (1.3) is assumed rather than working with sparsity in levels ($\Sigma_{s,M}$ from Def 1.4.1).

The flip test was first introduced in [15] to demonstrate the importance of the sparsity structure and we present variants of it here. The first flip test was designed to answer the following question: Is the success of subsampling techniques used in compressed sensing independent of the location of the coefficients to be recovered? In other words is sparsity the right model for compressed sensing?

The flip test can be described as follows Let $x \in \mathbb{C}^N$ be a vector, and $U \in \mathbb{C}^{N \times N}$ a measurement matrix. We then sample according to some pattern $\Omega \subseteq \{1, \dots, N\}$ with $|\Omega| = m$ and solve (1.16) for x , i.e. $\min \|z\|_1$ s.t $P_\Omega U z = P_\Omega U x$ to obtain a reconstruction $z = \alpha$. Now we flip x to obtain a vector x' with reverse entries, $x'_i = x_{N-i}$, $i = 1, \dots, N$

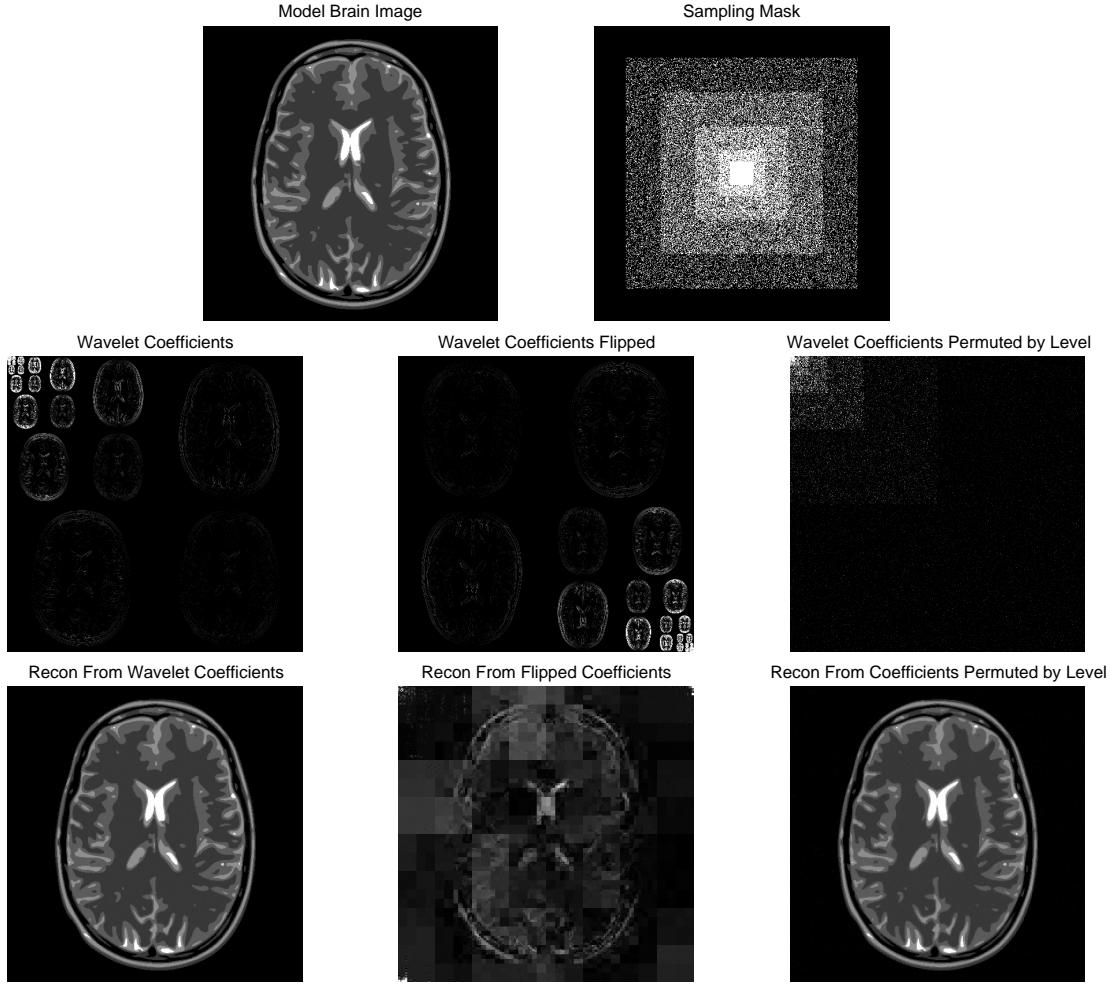


Figure 1.4: A standard flip test demonstrating the importance of the location of wavelet coefficients on the success of CS subsampling techniques. Here we decompose the Brain image into a two-dimensional separable Haar basis and solve for these coefficients through subsampling Fourier data of the corresponding image (Resolution= 512×512). We then compare the results to the same wavelet data but the order flipped versus randomly permuting the coefficients inside each wavelet level. After using CS to solve for the flipped/permuted coefficients, the data is then flipped/permuted back and then imaged.

and solve 1.16 for x' using the same U and Ω , i.e. $\min \|z\|_1$ s.t. $P_\Omega U z = P_\Omega U x'$. Assuming z to be a solution, then by flipping z we obtain a second reconstruction α' of the original vector x , where $\alpha'_i = z_{N-i}$. If the success of the sampling technique is independent of the structure of the coefficients, then we should have that α and α' are close. This is tested on the example of the GLPU analytical phantom from [36]. As Figure 1.4 suggests, this is not the case.

Another version of the flip test is to permute the coefficients in the wavelet levels that corresponds to the different scales. In other words, instead of flipping the coefficients

completely, the coefficients are permuted only within a certain level, but never across the levels. The result of this is visualised in Figure 1.4. Note that the results of the flip test presented here carry over to different subsampling schemes as well [37].

Conclusion of the flip tests:

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

It is clear from the flip test (see below) that sparsity is not the right model for the Fourier to wavelet case which is central to MRI and numerous other applications. In particular, sampling strategies that seek to recover sparse vectors will have to take an unrealistically large number of samples. This is also reflected in the theoretical guarantees of [35] which exhibit several additional logarithmic factors over traditional compressed sensing estimates.

1.7 Overview of the Thesis and its Contributions

This thesis has two key parts. Firstly, there is the work on asymptotic incoherence which looks at finding the optimal decay of the line and block coherences

$$\mu(R_N U), \mu(UR_N), \mu(\pi_N U), \mu(U\pi_N), \quad (1.42)$$

in various cases. Chapter 2 covers the basic theory and its application to some one-dimensional continuous CS problems, finishing with some theoretical discussion that contains several optimality results. Chapter 3 extends this work further with a focus on the analogous multidimensional cases. Chapter 4 covers the applications to Helium atom scattering (HAS) and how continuous CS is particularly relevant in this setting.

In this section we cover the main results from Chapters 2 & 3. An overview of Chapter 4 is provided at the beginning of that chapter.

1.7.1 One Dimensional Coherence Bounds

Our main results provide estimates for the precise convergence rates of $\mu(R_N U)$ and $\mu(UR_N)$ in the case of one-dimensional Fourier-wavelet and Fourier-polynomial bases spanning $L^2[0, 1]$. For $\epsilon \in (0, 1]$ fixed, the Fourier basis $B_f(\epsilon)$ is defined in (1.30). The standard wavelet basis B_w for a given Daubechies scaling function ϕ and wavelet ψ consists of functions of the form

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k). \quad (1.43)$$

Suppose $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ is the change of basis matrix formed by the pair of bases $(B_f(\epsilon), B_w)$ (see Definition 1.32 for a formal definition of U). Since $U : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ the bases must be indexed by \mathbb{N} which means we must enumerate the bases in some way using orderings. For $B_f(\epsilon)$ we enumerate with increasing frequency (a *frequency ordering*) and for B_w we order with j increasing (a *leveled ordering*).

To see how these orderings can come into, consider Figure 1.5. Here we see that the column line coherences (which correspond to $\mu(\pi_N U)$) roughly decay according to the absolute value of frequency. How do we make such observations precise?

Theorem 1.7.1 (1D Fourier-Wavelet Case). *Suppose $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$ where ρ is a frequency ordering and τ is a leveled ordering. Under these conditions (with $\epsilon \in I_{J,p}$, see Section 2.1.6), we have*

$$\mu(\pi_N U), \mu(U\pi_N) = \Theta(N^{-1}). \quad (1.44)$$

From Lemma 1.5.6 we also deduce that $\mu(R_N U), \mu(UR_N) = \Theta(N^{-1})$.

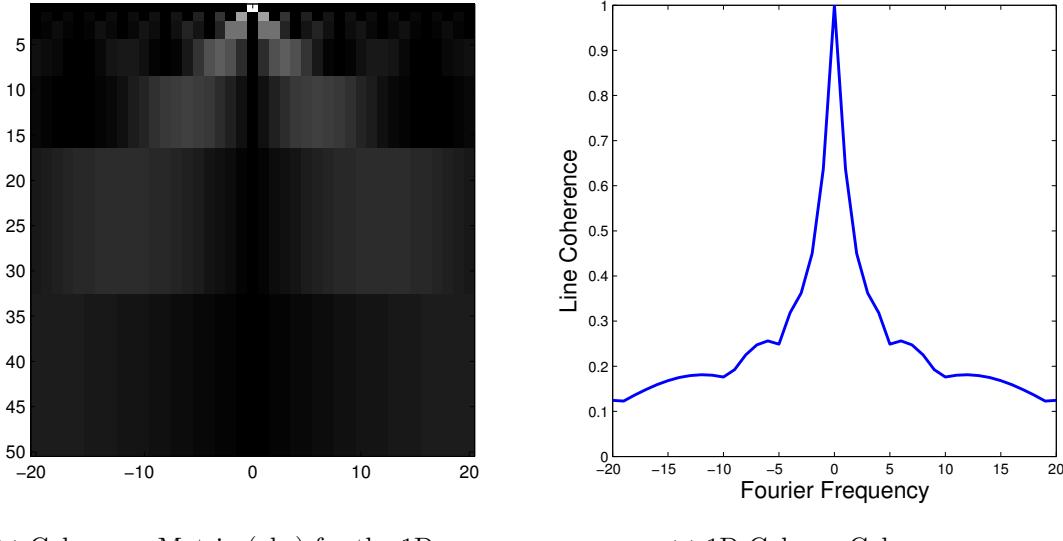
Remark 1.7.2. This is a combination of Theorems 2.2.1 and 2.2.5.

Next let $B_p = (\tilde{p}_n)_{n \in \mathbb{N}}$ denote the basis of L^2 -normalised Legendre polynomials on $[0, 1]$. Suppose $U \in \mathcal{B}(\ell^2(\mathbb{N}))$ now denotes the change of basis matrix formed by the pair of bases $(B_f(\epsilon), B_p)$ with a frequency ordering of $B_f(\epsilon)$ (B_p is already ordered by polynomial degree).

Theorem 1.7.3 (1D Fourier-Polynomial Case). *Suppose $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is the natural ordering and τ is a leveled ordering. In this case (given $\epsilon \in (0, 0.9]$), we have*

$$\mu(\pi_N U), \mu(U\pi_N) = \Theta(N^{-2/3}), \quad (1.45)$$

Figure 1.5: Fourier - Haar Case in One Dimension



In (b), the coherences are calculated by taking the maxima over the columns in (a), demonstrating decay that scales with frequency. In (a), white indicates larger absolute value.

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

Remark 1.7.4. This is a combination of Theorems 2.3.4 and 2.3.2.

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

1.7.2 Multidimensional Coherence Bounds

Tensor Cases

The multidimensional coherence results of the thesis can be broken down into two groups: one for tensor cases in general and one for the Fourier - separable wavelet case. In what follows $d \in \mathbb{N}$ denotes dimension. The definition of a tensor basis is given below.

Definition 1.7.5 (Tensor basis). Suppose that B is an orthonormal basis of some space $T \leq L^2(\mathbb{R})$ (i.e. T is a subspace $L^2(\mathbb{R})$) and we already have an ordering $\rho : \mathbb{N} \rightarrow B$.

Define $\rho^d : \mathbb{N}^d \rightarrow \bigotimes_{j=1}^d T \leq L^2(\mathbb{R}^d)$ by the formula ($m \in \mathbb{N}^d$)

$$\rho^d(m)(x) := \left(\bigotimes_{j=1}^d \rho(m_j) \right)(x) = \prod_{j=1}^d \rho(m_j)(x_j).$$

This gives a basis of $\bigotimes_{j=1}^d T \leq L^2(\mathbb{R}^d)$ because of the formula

$$\langle \rho^d(m), \rho^d(n) \rangle_{L^2(\mathbb{R}^d)} = \prod_{j=1}^d \langle \rho(m_j), \rho(n_j) \rangle_{L^2(\mathbb{R})}. \quad (1.46)$$

We call $B^d := (\rho^d(m))_{m \in \mathbb{N}^d}$ a ‘tensor basis’. The function ρ^d is said to be the ‘ d -dimensional indexing induced by ρ ’. Notice that ρ^d is not an ordering unless $d = 1$.

Example 1.7.6 (Multidimensional Fourier Basis). Recall the definition of χ_k from 1.30. The functions

$$\chi_k := \bigotimes_{j=1}^d \chi_{k_j}, \quad k \in \mathbb{Z}^d,$$

form the basis $B_f^d(\epsilon)$. It shall be convenient to identify $B_f^d(\epsilon)$ with \mathbb{Z}^d using the function

$$\lambda_d : B_f^d(\epsilon) \rightarrow \mathbb{Z}^d, \quad \lambda_d(\chi_k) := (\lambda(\chi_{k_1}), \dots, \lambda(\chi_{k_d})) = (k_1, \dots, k_d) = k. \quad (1.47)$$

Using tensor bases to construct multidimensional bases from a one dimensional basis is a standard trick. Suppose that we have a one-dimensional basis pair (B_1, B_2) and we know how the local coherence decays via results such as (1.44) and (1.45). Then the corresponding decay rates for the multidimensional tensor case (B_1^d, B_2^d) can be deduced using the technical Theorem 3.2.7.

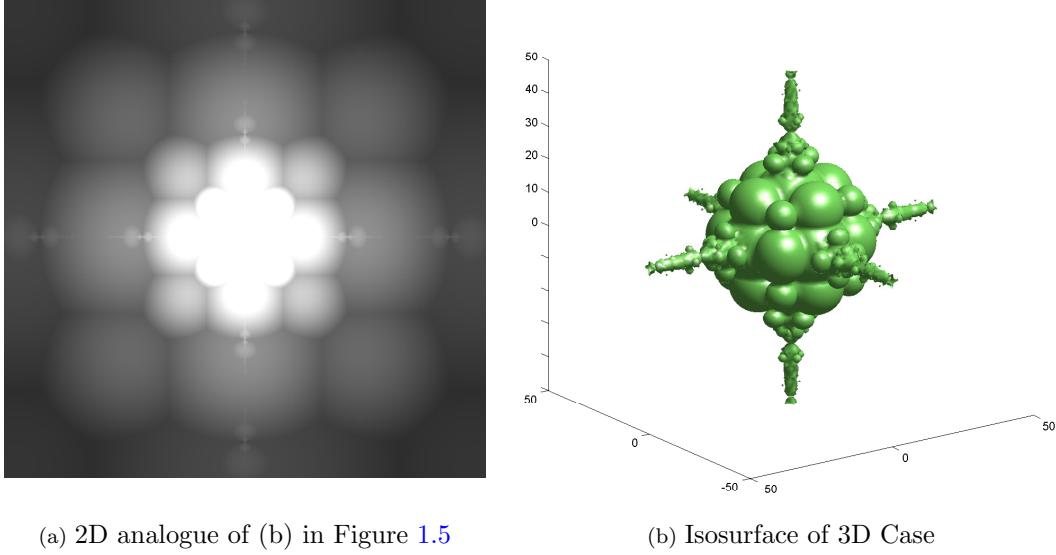
What makes this case both interesting and challenging is that the orderings that optimise the decay have to follow the *hyperbolic cross* very closely (c.f. 3.2.8). This can be a major drawback in many applications as this means taking very large frequency values (on a particular axis) relative to the reconstruction resolution .

Applying Theorem 3.2.7 to (1.44) and (1.45) gives us the following results:

Theorem 1.7.7 (Multi-Dimensional Fourier - Tensor Wavelet Case). Suppose that ρ and τ are as in Theorem 1.7.1. Let $U = [(B_f^d(\epsilon), \sigma^{-1} \circ \rho^d), (B_w^d, \sigma^{-1} \circ \tau^d)]$ where $\sigma : \mathbb{N}^d \rightarrow \mathbb{N}$ corresponds to the hyperbolic cross (see 3.2.8). Then, with $\epsilon \in I_{J,p}$,

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(\log^{d-1}(N+1) \cdot N^{-1}). \quad (1.48)$$

Figure 1.6: Fourier - Separable Haar Case in Two and Three Dimensions



In 2D this decay roughly matches that of the norm of the frequency as seen in (a). However in 3D there are hyperbolic spikes around the coordinate axes that lead to poor incoherence decay (see (b)) when using sampling patterns with rotational invariance or linear scaling. In (a), white indicates larger absolute value.

Consequently we also deduce that $\mu(R_N U), \mu(UR_N) = \Theta(\log^{d-1}(N+1) \cdot N^{-1})$.

Remark 1.7.8. This is a copy of Theorem 3.2.9.

Theorem 1.7.9 (Multi-Dimensional Fourier - Tensor Legendre Polynomial Case). Suppose that ρ and τ are as in Theorem 1.7.3. Let $U = [(B_f^d(\epsilon), \sigma^{-1} \circ \rho^d), (B_p^d, \sigma^{-1} \circ \tau^d)]$ where $\sigma : \mathbb{N}^d \rightarrow \mathbb{N}$ corresponds to the hyperbolic cross. Then, with $\epsilon \in I_{J,p}$,

$$\mu(\pi_N U), \mu(U\pi_N) = \Theta((\log^{d-1}(N) \cdot N^{-1})^{2/3}). \quad (1.49)$$

Consequently we also deduce that $\mu(R_N U), \mu(UR_N) = \Theta((\log^{d-1}(N) \cdot N^{-1})^{2/3})$.

Remark 1.7.10. This is a copy of Theorem 3.2.10.

The additional logarithmic factors in the tensor cases here demonstrates the typical problems associated with dimensionality.

Separable Wavelet Cases

Although using tensor polynomials (multinomials) is relatively commonplace, it is uncommon to see tensor wavelets used in practice. Instead it is separable wavelets [21] that is used in most numerical wavelet transforms as they have clear resolution levels with well-controlled support. This case is both the most interesting and complex case of study that is known to date. The definition of a separable wavelet basis B_{sep}^d is provided in Section 3.3.

Since the separable wavelet structure still has resolutions levels, ordering this basis according to these levels is relatively straightforward (c.f. 3.3.3). The primary challenge of handling the Fourier - separable wavelet case is the question of how to order the frequency basis. Following the one dimensional case one can try and order the frequencies in \mathbb{Z}^d by some notion of size:

Definition 1.7.11. *Any ordering $\rho : \mathbb{N} \rightarrow B_f^d(\epsilon)$ that is consistent with the function $F(\chi_k) = \|k\|$, where $\|\cdot\|$ is a norm on \mathbb{R}^d , is called a norm ordering.*

This seems like a natural extension of ordering by size in one dimension however, it turns out the full picture is not quite that simple:

Theorem 1.7.12 (Fourier - Separable Wavelet Case for Norm Orderings). *Suppose $U = [(B_f^d(\epsilon), \rho), (B_{\text{sep}}^d, \tau)]$ where ρ is a norm ordering and τ is a leveled ordering. Assuming that the wavelet generating B_w is sufficiently smooth and $\epsilon \in I_{J,p}$, we have*

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1}). \quad (1.50)$$

Consequently it follows that $\mu(R_N U), \mu(U R_N) = \Theta(N^{-1})$.

Remark 1.7.13. *This is a corollary of Theorems 3.3.11 and 3.3.5.*

The key difference of this with Theorem 1.7.1 is the assumption on wavelet smoothness. In 2D, this condition can be removed but in three dimensions or more this is necessary:

Lemma 1.7.14. *Let B_{sep}^3 be a separable 3D Haar wavelet basis, then there exists $C > 0$ such that*

$$\mu(R_N U) \geq CN^{-2/3}. \quad (1.51)$$

Remark 1.7.15. *This follows from Example 3.3.13.*

Faced with this problem, we cannot use norm orderings to tackle every multidimensional case. Instead a class of orderings have to be introduced, called *semihyperbolic orderings*:

Definition 1.7.16. Define, for $r, d \in \mathbb{N}, r \leq d$ the function

$$H_{d,r}(\chi_k) := \max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{j=1}^r \max(|k_{i_j}|, 1), \quad k \in \mathbb{Z}^d.$$

Then we say an ordering $\sigma : \mathbb{N} \rightarrow B_f^d$ is ‘semi-hyperbolic of order r in d dimensions’ if it is consistent with $H_{d,r}$.

This turns out to cover both norm orderings (case $r = 1$) and the orderings based on the hyperbolic cross used in the tensor case (case $r = d$). With this definition, the smoothness requirement in Theorem 1.7.12 can be removed:

Theorem 1.7.17 (Fourier - Separable Wavelet Case for Semi-Hyperbolic Orderings). Suppose $U = [(B_f^d(\epsilon), \rho), (B_{sep}^d, \tau)]$ where ρ is a semihyperbolic ordering of order $d - 1$ and τ is a leveled ordering. Then, assuming $\epsilon \in I_{J,p}$,

$$\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1}). \quad (1.52)$$

Consequently it follows that $\mu(R_N U), \mu(U R_N) = \Theta(N^{-1})$.

Remark 1.7.18. This follows from Theorems 3.3.19 and 3.3.5.

This is a significant improvement over the tensor case, removing the additional logarithmic factors.

1.7.3 Optimality Results

An important question is whether any of the above results are actually the best possible; is it possible to use different orderings to improve upon these results and achieve faster decay rates? In this thesis a theory of optimal orderings, introduced by the author in [1] is introduced to answer this question. Sections 2.5 and 2.6 covers the following results:

Theorem 1.7.19 (Optimality). For the 1D Fourier-wavelet case $(B_f(\epsilon), B_w)$, none of the following decay rates can be achieved

$$\mu(R_N U) = o(N^{-1}), \quad \mu(U R_N) = o(N^{-1}). \quad (1.53)$$

For the 1D Fourier-polynomial case ($B_f(\epsilon)$, B_p), the following decay rates are impossible no matter what orderings of the bases are used:

$$\mu(R_N U) = o(N^{-2/3}), \quad \mu(UR_N) = o(N^{-2/3}). \quad (1.54)$$

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

$$\sum_{N \in \mathbb{N}} \mu(R_N U) < \infty, \quad (1.55)$$

must hold in general. Therefore $\mu(R_N U) = o(N^{-\alpha})$ is impossible for $\alpha > 1$ showing that the Fourier-wavelet decay (1.44) attains the fastest theoretically possible decay rate as a power law. Furthermore it is shown that, as a statement for all isometries $U \in \mathcal{B}(\ell^2(\mathbb{N}))$, (1.55) cannot be improved upon.

1.7.4 Other Results

There are other smaller results in this thesis that we have not covered in this overview (such as optimality of linear orderings implying wavelet smoothness in Section 3.3.7 or hierarchies of different orderings in Section 3.3.8) but these are left to be discussed in the conclusion and their relevant sections.

1.7.5 How this Work Developed

This is a good point to mention how my thesis began and what was the state of development of local coherence theory when I joined the AFHA group as a PhD student. When I joined the work the breaking coherence paper [15] was already well developed and the need to study local coherence decay (known as *asymptotic incoherence* to the group) was already identified. Progress on the one dimensional wavelet and polynomial cases had already been made in the sparsity directions. In particular the upper bounds (i.e. \mathcal{O} not Θ) in the sparsity direction ($\mu(U\pi_N)$) for (1.44) and (1.45) were already known. My original role was to extend these two bounds to the sampling direction ($\mu(\pi_N U)$) and show that the results were tight.

At the start of the work the orderings of each basis was essentially kept fixed. After the one-dimensional cases were relatively settled, tentative attempts at the 2D Fourier - Tensor Wavelet case were made. Eventually the need to use the hyperbolic cross

to optimise the coherence decay in this case forced me to place greater and greater importance on the orderings. This lead to the development of the notion of an optimal ordering. From this the theoretical limits on optimal decay were established and the 2D case was complete for both tensor and separable wavelet cases.

The next stumbling block was tackling the 3D Fourier - Separable Haar Wavelet case. Since linear orderings worked for all 2D separable wavelet cases it was suspected that a wavelet independent approach could be used in general but this was quickly proved wrong by Example 3.3.13. Around this time the d D tensor case was complete and so the hyperbolic cross was tried in this case. It was an improvement (see Proposition 3.3.6) however it did not achieve the theoretically optimal rate. Finally it clicked that a cross between linear and hyperbolic orderings were necessary and the notion of semihyperbolic orderings finally finished off the separable wavelet cases in general.

Chapter 2

One Dimensional Cases & Optimality Results

In this chapter we cover the one dimensional Fourier - Wavelet and Fourier - Polynomial cases in Sections 2.2 and 2.3 with accompanying numerical results in Section 2.4. Afterwards notions of optimality are covered in Section 2.5 with a discussion of more general theoretical limits left to its own section in Section 2.6. A discussion on the definition of optimality is presented in Section 2.7. Finally, a derivation of the $s \log(N)$ sampling rate for the Fourier-Daubechies wavelet case is presented in Section 2.8.

2.1 Bases and Orderings

2.1.1 Fourier Basis

We recall the definition of the Fourier basis $B_f(\epsilon)$ from (1.30).

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

$$\rho(m)(x) = \sqrt{\epsilon} \exp(2\pi i \epsilon \cdot \lambda \circ \rho(m)x) \cdot \mathbb{1}_{[0,\epsilon^{-1}]}(x), \quad \forall m \in \mathbb{N}.$$

Definition 1.5.3 says that an ordering ρ of $B_f(\epsilon)$ is a frequency ordering if and only if the function $|\lambda \circ \rho|$ is nondecreasing. Therefore ρ is a frequency ordering if and only if we have $\{\lambda \circ \rho(2n), \lambda \circ \rho(2n+1)\} = \{+n, -n\}$ for $n \in \mathbb{N}$ and $\lambda \circ \rho(1) = 0$ and consequently $|\lambda \circ \rho(m)| = \lceil (m-1)/2 \rceil$.

2.1.2 Legendre Bases

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 on $[0, 1]$ are defined by $\tilde{p}_n(x) = \sqrt{2n-1} \cdot p_N(2x-1)$ and we write $B_p := (\tilde{p}_n)_{n=1}^\infty$. The basis B_p is already ordered; call this the *natural ordering*.

2.1.3 Standard Wavelets

Note: A full construction of Daubechies basis will not be presented here; it will be assumed that the mother wavelet and scaling function has already been constructed. Those seeking for a complete description should consult [26] or [22].

Take a Daubechies wavelet ψ with $p \in \mathbb{N}$ vanishing moments and its corresponding scaling function ϕ in $L^2(\mathbb{R})$ where

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

The scaled and shifted versions of ψ, ϕ are denoted by

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

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

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

where W_j here is the orthogonal complement of V_j in V_{j+1} . For a fixed $J \in \mathbb{N}$ we define the set¹

$$B_w := \left\{ \phi_{J,k}, \psi_{j,k} : \begin{array}{l} \text{Supp}(\phi_{J,k}) \cap (0, 1) \neq \emptyset, \\ \text{Supp}(\psi_{j,k}) \cap (0, 1) \neq \emptyset, \\ j \in \mathbb{N}, j \geq J, k \in \mathbb{Z} \end{array} \right\}, \quad (2.1)$$

Let ρ be an ordering of B_w . Since $L^2(\mathbb{R}) = \overline{V_J \oplus \bigoplus_{j=J}^\infty W_j}$ for all $f \in L^2(\mathbb{R})$ we have

¹'w' here stands for 'wavelet'.

$$f(x) = \sum_{k \in \mathbb{Z}} \left(\langle f, \phi_{J,k} \rangle \phi_{J,k}(x) + \sum_{j=J}^{+\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k}(x) \right). \quad (2.2)$$

and so for f with $\text{supp}(f) \subseteq [0, 1]$ we have

$$f = \sum_{n=1}^{\infty} c_n \rho(n) \quad \text{for some } (c_n)_{n \in \mathbb{N}} \in \ell^2(\mathbb{N}). \quad (2.3)$$

Definition 2.1.1 (Leveled ordering (standard wavelets)). Define $F_w : B_w \rightarrow \mathbb{R}$ by

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

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

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

2.1.4 Periodic Wavelets

Since the Fourier series approximation 1.10 approximates the periodised version of the function f it is natural (and in fact desirable) to search for a wavelet expansion that also reconstructs the periodised version of f . As we work over an interval, it would be convenient not to have to use the standard wavelet approach as their supports sometimes crossover the interval boundary.

For a function $g \in L^2(\mathbb{R})$ let $\tilde{g}(x) = \sum_{k \in \mathbb{Z}} g(x - k)$ denote the $[0, 1]$ -periodised version² of g . Then following on from 2.2 we have that

$$\tilde{f}(x) = \sum_{k=1}^{2^J} \langle \tilde{f}, \tilde{\phi}_{J,k} \rangle \tilde{\phi}_{J,k}(x) + \sum_{j=J}^{+\infty} \sum_{k=1}^{2^j} \langle \tilde{f}, \tilde{\psi}_{j,k} \rangle \tilde{\psi}_{j,k}(x). \quad (2.4)$$

Consequently we deduce that the following collection of functions forms a basis for

²We assume that the function is either compactly supported or decays sufficiently fast for this periodisation to be well-defined.

$L^2[0, 1]$:

$$B_{\text{pw}} = \{\tilde{\phi}_{J,k} : k = 1, \dots, 2^J\} \cup \bigcup_{j=J}^{\infty} \{\tilde{\psi}_{j,k} : k = 1, \dots, 2^j\}. \quad (2.5)$$

Two examples of periodic wavelet bases are shown in Figure 4.7 along with an application to a single exponential peak.

2.1.5 Boundary Wavelets

An alternative interval-based approach is to use boundary wavelets [21, Section 7.5.3], originally developed in [38]. The basis functions all have support contained within $[0, 1]$, while still spanning $L^2[0, 1]$. Furthermore, the boundary wavelet basis retains the ability to reconstruct polynomials of order up to $p - 1$ from the corresponding standard wavelet basis (which does not hold for the periodic wavelet basis). We shall not go into great detail here but we will outline the construction; we take, along with a Daubechies wavelet ψ and corresponding scaling function ϕ with $\text{Supp}(\psi) = \text{Supp}(\phi) = [-p+1, p]$, boundary scaling functions and wavelets (using the same notation as in [21])

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

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

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

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

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

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

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

Boundary wavelets do not approximate the periodic version of a function which, while not a problem for truly interval-based problems, is a disadvantage for Fourier-transform

based applications where the Fourier series approximation enforces a periodised target reconstruction function.

2.1.6 A Few Comments on the Wavelet Bases

Both the periodic (B_{pw}) and boundary wavelet (B_{bw}) bases also have corresponding leveled orderings following the same definition as 2.1.1.

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

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

If B_w is replaced by B_{bw} , we only require $\epsilon \leq 1$, since every function in B_{bw} has support contained in $[0, 1]$. For the periodic wavelet basis B_{pw} , only $\epsilon = 1$ make real sense, otherwise periodised versions of f in the Fourier series expansion (1.10) and (2.4) do not agree. For the rest of this thesis, we shall assume these constraints on ϵ hold.

2.2 1D Fourier-Wavelet Case

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

$$\begin{aligned} U_{m,n} &= \langle \tau(n), \rho(m) \rangle = \int_{\mathbb{R}} \sqrt{\epsilon} \exp(-2\pi i \epsilon x \cdot \lambda \circ \rho(m)) \cdot \tau(n)(x) \, dx \\ &= \sqrt{\epsilon} \mathcal{F}\tau(n)(\epsilon \cdot \lambda \circ \rho(m)), \end{aligned} \tag{2.6}$$

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

$$\begin{aligned} \mathcal{F}\phi_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\phi(2^{-j} \omega), & \mathcal{F}\psi_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\psi(2^{-j} \omega), \\ \mathcal{F}\tilde{\phi}_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\phi(2^{-j} \omega), & \mathcal{F}\tilde{\psi}_{j,k}(\omega) &= e^{-2\pi i 2^{-j} k \omega} 2^{-j/2} \mathcal{F}\psi(2^{-j} \omega), \\ \mathcal{F}\psi_{j,n}^{\text{left}}(\omega) &= 2^{-j/2} \mathcal{F}\psi_n^{\text{left}}(2^{-j} \omega), & \mathcal{F}\psi_{j,n}^{\text{right}}(\omega) &= 2^{-j/2} e^{-2\pi i} \mathcal{F}\psi_n^{\text{right}}(2^{-j} \omega). \end{aligned} \tag{2.7}$$

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

Theorem 2.2.1. *Let $U = [(B_f(\epsilon), \rho), (B_{\text{pw}}, \tau)]$ where τ is a leveled ordering of a periodic*

wavelet basis. Then for $\epsilon = 1$ we have

$$\mu(U\pi_N) = \Theta(N^{-1}). \quad (2.8)$$

Furthermore, suppose instead $U = [(B_f(\epsilon), \rho), (B_w, \tau)]$ where τ is a leveled ordering of a standard wavelet basis. Then (2.8) holds for every $\epsilon \in I_{J,p}$. Likewise for the boundary wavelet basis B_{pw} for every $\epsilon \in (0, 1]$.

Below we give the proof for periodic wavelets as it is the cleanest of the arguments. The other two cases are very similar and their proof can be found in [1].

Proof: (Periodic Wavelets). By equation (2.6) we know that (since $\lambda \circ \rho : \mathbb{N} \rightarrow \mathbb{Z}$ is bijective)

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

where $j(N) \in \mathbb{N}$ denotes the scaling level of $\tau(N) \in B_{pw}$. Notice that for a leveled ordering of B_{pw} , the functions belonging to the scaling level V_J come first, and there are of 2^J of these functions. Therefore, for $N \leq 2^J$ we have that, by (2.7),

$$\mu(U\pi_N) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-J} |\mathcal{F}\phi(2^{-J}\epsilon m)|^2. \quad (2.9)$$

Furthermore, for $\tau(N)$ with $N > 2^J$, we have that, by (2.7),

$$\mu(U\pi_N) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-j(N)} |\mathcal{F}\psi(2^{-j(N)}\epsilon m)|^2. \quad (2.10)$$

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

$$\frac{\sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(\epsilon 2^{-j(N)} m)|^2}{\sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2} \longrightarrow 1 \quad \text{as } N \rightarrow \infty. \quad (2.11)$$

We are therefore left with handling the $2^{-j(N)}$ term, which means estimating $j(N)$ as $N \rightarrow \infty$.

We can bound $j(N)$ by counting how many wavelet functions are in each scaling level; for the j th level there are 2^j wavelet functions. Therefore, the following inequality

holds for $N > 2^J$:

$$2^{j(N)} \leq N \leq 2^{j(N)+1}. \quad (2.12)$$

consequently we deduce that $2^{-j(N)} = \Theta(N^{-1})$. Combining this with (2.11) gives us $\mu(U\pi_N) = \Theta(N^{-1})$ as required. \square

Remark 2.2.2. In [1] the stronger result (for standard/boundary wavelets) was proved:

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

$$\frac{\epsilon \cdot C_1}{N} \leq \mu(U\pi_N) \leq \frac{\epsilon \cdot C_2}{N}. \quad (2.13)$$

Furthermore, suppose instead $U = [(B_f(\epsilon), \rho), (B_{bw}, \tau)]$ where τ is a leveled ordering of a boundary wavelet basis. Then there are constants $C_1, C_2 > 0$, dependent on the choice of wavelet, such that for all $\epsilon \in (0, 1]$ and $N \in \mathbb{N}$, (2.13) holds.

In order to bound the line coherences in the sampling direction ($\mu(\pi_N U)$), we need the following condition on our scaling function / wavelet; there exists a constant $K > 0$ s.t. $\forall \omega \in \mathbb{R} \setminus \{0\}$,

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \quad (2.14)$$

This condition holds for all Daubechies wavelets (see the proof of Proposition 4.7 in [39]), in fact it even holds if we change the power of ω from $1/2$ to 1 .

Lemma 2.2.3. Let ϕ be a Daubechies scaling function, with corresponding mother wavelet ψ . Then, along with (2.14), we also have

$$|\mathcal{F}\psi(\omega)| \leq \frac{K}{|\omega|^{1/2}}. \quad (2.15)$$

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

$$|\mathcal{F}\phi_n^{left}(\omega)|, |\mathcal{F}\phi_n^{right}(\omega)|, |\mathcal{F}\psi_n^{left}(\omega)|, |\mathcal{F}\psi_n^{right}(\omega)| \leq \frac{K}{|\omega|^{1/2}}, \quad (2.16)$$

along with (2.14) and (2.15). In fact (2.15) and (2.16) hold with the powers of $1/2$ replaced by 1 .

Proof. We notice that if (2.14) holds then we can use the equation (from (2.14) in [22])

$$\mathcal{F}\psi(2\omega) = \exp(2i\pi\omega) \cdot \nu(2\omega) \cdot m_0(\omega + 1/2) \cdot \mathcal{F}\phi(\omega), \quad (2.17)$$

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 1³. 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 [22]) we can show that (2.14) 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 2.14 is closed under finite linear combinations. Next observe that if we prove an estimate of the form 2.14 for the functions (see page 71 of [38])

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

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

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

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

1. $|\mathcal{F}\mathbf{1}_{[0,2p-1]}(\omega)| = |(\exp(-2\pi i(2p-1)\omega) - 1) \cdot (2\pi i\omega)^{-1}| \leq C_1 \cdot (|\omega| + 1)^{-1}$ for some constant $C_1 > 0$.
2. Excluding the Haar wavelet, for every Daubechies wavelet there exists constants $\alpha, C_2 > 0$ such that $|\mathcal{F}\phi(\omega)| \leq C_2 \cdot (|\omega| + 1)^{-1-\alpha}$ (see the proof of Proposition 4.7 in [39]).

³The equation here is not identical to that of the reference because of our choice of definition of the Fourier transform.

We now claim that if two functions f, g satisfy

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

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

$$\begin{aligned} |f * g(\omega)| \cdot |\omega| &\leq C_1 C_2 \int_{\mathbb{R}} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \\ &\leq C_1 C_2 \left(\int_{-\infty}^{\omega/2} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \right. \\ &\quad \left. + \int_{\omega/2}^{+\infty} \frac{|\omega|}{(|u| + 1)(|\omega - u| + 1)^{1+\alpha}} du \right), \end{aligned} \quad (2.18)$$

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

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

Lemma 2.2.4. *For any compactly supported wavelet ψ with scaling function $\phi \in L^1(\mathbb{R})$ there exists an $N \in \mathbb{N}$ such that for all $q \geq N$, ($q \in \mathbb{N}$) we have*

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

Proof. We recall from equation (2.17) that

$$|\mathcal{F}\psi(2\omega)| = |m_0(\omega + 1/2)| \cdot |\mathcal{F}\phi(\omega)|. \quad (2.19)$$

Furthermore, we also know that $|\mathcal{F}\phi(0)| = 1$ and $m_0(1/2) = 0$ [22]⁴. However, since ϕ is compactly supported, m_0 is a non-zero trigonometric polynomial and so it follows that this zero at $1/2$ is isolated. Therefore, since $\mathcal{F}\phi$ is continuous, we deduce that (2.19) is nonzero when $\omega > 0$ is sufficiently small. \square

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

Theorem 2.2.5. *Let $U = [(B_f(\epsilon), \rho), (B_{pw}, \tau)]$ where ρ is a frequency ordering of the Fourier basis. Then for $\epsilon = 1$ we have*

$$\mu(\pi_N U) = \Theta(N^{-1}). \quad (2.20)$$

If we replace B_{pw} by B_w in the above setup, then (2.20) also holds for each $\epsilon \in I_{J,p}$. Likewise for B_{bw} and each $\epsilon \in (0, 1]$.

Proof: (Periodic/Standard Wavelets). **Upper Bound:** Since ρ is a frequency ordering if $m = 1$ then $\lambda \circ \rho(m) = 0$ and since $|\mathcal{F}\phi(0)| = 1$, $\mathcal{F}\psi(0) = 0$ (see (2.19) and the line below it) then we have $\mu(\pi_1 U) = \epsilon 2^{-J}$. Next let $m \geq 1$. Observe that the estimate (2.14) is strong enough to bound the finitely many $\phi_{J,k}$ terms as required since

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

where we used that ρ is a frequency ordering in the last step. Therefore we are left with the terms involving the shifts and dilations of ψ . This is also a straightforward consequence of (2.14) since we have

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

This gives the global bound for $m \geq 2$ (uniform in n)

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

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

⁴See Section 2 Theorem 1.7 and Equation (3.1) in the reference.

Lower Bound: 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 2.2.4, 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 a frequency ordering. Recall that by Lemma 2.2.4 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. This completes the lower bound. \square

Remark 2.2.6. Likewise the following stronger result was also proved by the author in [1]:

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

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

Furthermore, there is a constant $C_2 > 0$ such that for all $\epsilon \in I_{J,p}$ and $N \geq 1 + 2^J \epsilon^{-1}$ 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 $\epsilon \in (0, 1]$.

2.3 1D Fourier-Polynomial Case

Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is frequency ordering and τ the natural ordering of the Legendre polynomials.

Outline of Argument: The key fact we use in this section is that the entries of U

are directly related to Bessel functions (see (2.25)):

$$\begin{aligned} |U_{m,n}| &= 2\sqrt{\epsilon(n - 1/2)} \cdot |j_{n-1}(2\pi\epsilon\lambda \circ \rho(m))| \\ &= \frac{\sqrt{n - 1/2}}{\sqrt{\lambda \circ \rho(m)}} \cdot |J_{n-1/2}(2\pi\epsilon\lambda \circ \rho(m))|, \quad (m \neq 1), \end{aligned} \tag{2.21}$$

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

Before we start formally covering the line coherence estimates for these two bases we shall first need to prove a preliminary result.

Lemma 2.3.1. *Let J_n denote the n th Bessel function of the first kind and let $j'_{n,k}$ denote the k th non-negative root of J'_n . Furthermore, Let j_n denote the n th spherical Bessel function of the first kind and let $a'_{n,k}$ denote the k th non-negative root of j'_n . Then if $n \geq 1$ we have*

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

Proof. The result for J_n follows from the arguments given in [40, 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 [41, Eqn. (10.1.2)]

$$j_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m 2^{n+1} (n+m+1)! x^{n+2m}}{m! (2(n+m+1))!}. \tag{2.22}$$

Second, we shall use the fact that j_n is a solution to the following differential equation [41, Eqn. (10.1.1)]

$$x^2 j_n(x)'' + 2x j_n'(x) + (x^2 - n(n+1)) j_n(x) = 0. \tag{2.23}$$

We first observe that by (2.22), $|j_n(-x)| = |j_n(x)|$, $\forall x \in \mathbb{R}$ and so we need only consider

$\sup_{x \in [0, +\infty)} |j_n(x)|$. (2.23) can be rephrased as

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

Therefore, noting that by (2.22), $j_n(x) > 0$ 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

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

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

$$\begin{aligned} \Lambda'_n(x) &= 2j'_n(x)j_n(x) + \frac{2xj'^2_n(x) + 2x^2j'_n(x)j''_n(x)}{x^2 - n(n+1)} - \frac{2x^3j'^2_n(x)}{(x^2 - n(n+1))^2} \\ &= \frac{2j'_n(x)(j_n(x)(x^2 - n(n+1)) + xj'_n(x) + x^2j''_n(x))}{x^2 - n(n+1)} - \frac{2x^3j'^2_n(x)}{(x^2 - n(n+1))^2} \quad (2.24) \\ &= -\frac{2xj'^2_n(x)}{x^2 - n(n+1)} - \frac{2x^3j'^2_n(x)}{(x^2 - n(n+1))^2} < 0. \quad (\text{using (2.23)}) \end{aligned}$$

This tells that $|j_n(a'_{n,1})| > |j_n(a'_{n,2})| > |j_n(a'_{n,3})| \dots$. To finish the proof we notice that by (2.22), $j_n(0) = 0$ for $n \geq 1$ and furthermore, by [41, Eqn. (10.1.14)],

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

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

Theorem 2.3.2. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where τ is the natural ordering of the polynomial basis. Then for all $\epsilon \in (0, 0.9]$ we have*

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

Proof. **Upper Bound:** First notice that if $\epsilon' = \epsilon/2$

$$\begin{aligned}
U_{m,n} &= \langle \rho(m), \tilde{p}_n \rangle_{L^2([0,1])} \\
&= \sqrt{\epsilon'} \cdot \sqrt{n-1/2} \int_{-1}^1 e^{2\pi i \lambda \circ \rho(m) \epsilon' t} p_n(t) dt \\
&= i^{n-1} 2\sqrt{\epsilon'(n-1/2)} \cdot j_{n-1}(2\pi \epsilon' \lambda \circ \rho(m)) \\
&= i^{n-1} \frac{\sqrt{n-1/2}}{\sqrt{\lambda \circ \rho(m)}} \cdot J_{n-1/2}(2\pi \epsilon' \lambda \circ \rho(m)), \quad (m \neq 1)
\end{aligned} \tag{2.25}$$

where on the third line we have used [41, Eqn. (10.1.14)] and on the fourth line we have used the following formula connecting the spherical Bessel function to the standard Bessel function:

$$j_n(z) = \sqrt{\frac{\pi}{2z}} J_{n+1/2}(z). \tag{2.26}$$

Therefore, we find

$$\mu(U\pi_N) \leq 4\epsilon'(N-1/2) \sup_{t \in \mathbb{R}} j_{N-1}^2(t). \tag{2.27}$$

We therefore need to estimate $\sup_{t \in \mathbb{R}} |j_n(t)|$. By Lemma 2.3.1, we know that

$$\sup_{t \in \mathbb{R}} |j_n(t)| = |j_n(a'_{n,1})|, \quad n \geq 1,$$

where $a'_{n,1}$ denotes the first positive root of j'_n .

Thus, we only need to have estimates for $|j_n(a'_{n,1})|$. But we also know [41, Eqn. (10.1.61)], that the following asymptotic expansion holds

$$j_n(a'_{n,1}) \sim \gamma(n+1/2)^{-5/6} + \mathcal{O}((n+1/2)^{-3/2}), \tag{2.28}$$

for some positive constant $1/2 < \gamma < 1$. Therefore 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 (2.27) we get the upper bound

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

Therefore the upper bound is complete for the case $N > N'$ (and notice that N' is

independent of ϵ'). However since $\sup_{x \in \mathbb{R}} |j_N(x)| < \infty$ for every N we can use (2.27) to cover the case $N \leq N'$, completing the upper bound.

Lower Bound: We focus on the following equation taken from (2.25)

$$|U_{m,n}| = \frac{\sqrt{n-1/2}}{\sqrt{|\lambda \circ \rho(m)|}} \cdot |J_{n-1/2}(2\pi\epsilon'\lambda \circ \rho(m))|, \quad (m \neq 1). \quad (2.29)$$

Let $j'_{\nu,1}$ denote the first positive zero of J'_ν . From [41, Eqns. (9.5.16), (9.5.20)], we have the asymptotic estimates

$$j'_{\nu,1} \sim \nu + \zeta \nu^{1/3} + \mathcal{O}(\nu^{-1/3}), \quad (2.30)$$

$$J(j'_{\nu,1}) \sim \kappa \cdot \nu^{-1/3} + \mathcal{O}(\nu^{-1}), \quad (2.31)$$

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

1. $2J'_\nu(x) = J_{\nu-1}(x) - J_{\nu+1}(x)$, [40, p. 45],
2. $\sup_{x \in \mathbb{R}} |J_\nu(x)| = |J_\nu(j'_{\nu,1})|$, using Lemma 2.3.1 .

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 (2.32) below.

By a triangle inequality $|J_{n-1/2}(k_n)| \geq |J_{n-1/2}(j'_{n-1/2,1})| - |J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})|$ and we bound the latter term by using integrals⁵:

$$\begin{aligned} |J_{n-1/2}(k_n) - J_{n-1/2}(j'_{n-1/2,1})| &= \left| \int_{j'_{n-1/2,1}}^{k_n} J'_{n-1/2,1}(t) dt \right| \\ &= \left| \int_{j'_{n-1/2,1}}^{k_n} \int_{j'_{n-1/2,1}}^t J''_{n-1/2,1}(u) du dt \right| \\ &\leq \left| \int_{j'_{n-1/2,1}}^{k_n} \int_{j'_{n-1/2,1}}^t |J_{n-1/2}(j'_{n-1/2,1})| du dt \right| \\ &\leq \frac{|j'_{n-1/2,1} - k_n|^2}{2} \cdot |J_{n-1/2}(j'_{n-1/2,1})| \\ &\leq \frac{(\pi\epsilon')^2}{2} \cdot |J_{n-1/2}(j'_{n-1/2,1})|. \end{aligned} \quad (2.32)$$

⁵The use of the second integral is valid since $J'_{n-1/2,1}(t) = J'_{n-1/2,1}(t) - J'_{n-1/2,1}(j'_{n-1/2,1})$ by the definition of $j'_{n-1/2,1}$.

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 (2.32) becomes

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

and therefore we deduce

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

Combining this inequality with (2.30), (2.31) gives us the following bound:

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

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

$$\frac{\sqrt{n-1/2}}{\sqrt{k_n}} \cdot |J_{n-1/2}(k_n)| \geq C n^{-1/3}. \quad (2.33)$$

Therefore, given $n \geq M$, let $m(n) \in \mathbb{N}$ be such that $2\pi\epsilon'\lambda \circ \rho(m) = k_n$. Then by (2.33) and (2.29) we have

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

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

For $N \leq M$ we observe that from (2.25)

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

As before we observe that since $j_{N-1}(x) \rightarrow 0$ as $x \rightarrow \infty$, the supremum

$$\sup_{m \in \mathbb{Z}} |j_{N-1}(2\pi\epsilon'm)|,$$

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

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

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

This argument actually shows the following stronger (but more technical) result:

Theorem 2.3.3. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where τ is the natural ordering of the polynomial basis. Then there are constants $C_1, C_2 > 0$ such that for all $\epsilon \in (0, 0.9]$ and $N \in \mathbb{N}$,*

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

Now to cover the line coherence in the sampling direction:

Theorem 2.3.4. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is a frequency ordering of the Fourier basis. The for $\epsilon \in (0, 1]$ we have*

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

Proof. Upper Bound: Without loss of generality we can assume τ is the natural ordering of B_p . Recall that from (2.25) we have

$$|U_{m,n}|^2 = \frac{n-1/2}{|\lambda \circ \rho(m)|} J_{n-1/2}^2(2\pi\epsilon' \lambda \circ \rho(m)) \quad (2.34)$$

$$= 4\epsilon'(n-1/2) j_{n-1}^2(2\pi\epsilon' \lambda \circ \rho(m)). \quad (2.35)$$

We shall first derive two useful bounds; notice that if we apply (2.28) to (2.35) then we get the bound, for some constant $\beta > 0$,

$$|U_{m,n}|^2 \leq 4\epsilon'(n-1/2) \cdot (\beta(n-1/2)^{-5/6})^2 \leq 4\epsilon'\beta^2(n-1/2)^{-2/3}. \quad (2.36)$$

Secondly we shall use the following inequality from [42]

$$|J_\nu(x)| \leq b\nu^{-1/3} \quad \nu > 0, \quad x \in \mathbb{R}, \quad (2.37)$$

where $b > 0$ is some constant. Applying this to (2.34) gives the bound

$$|U_{m,n}|^2 \leq \frac{n - 1/2}{|\lambda \circ \rho(m)|} (b(n - 1/2)^{-1/3})^2 \leq \frac{b^2(n - 1/2)^{1/3}}{|\lambda \circ \rho(m)|}. \quad (2.38)$$

Recall that our goal is to estimate $|U_{m,n}|$ uniformly in n as $m \rightarrow \infty$. We first apply the case $n - 1/2 \geq \epsilon' |\lambda \circ \rho(m)|$ to (2.36) 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 (2.38) 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 $\epsilon' \in (0, 1]$. If $m = 0$, i.e. $\lambda \circ \rho(m) = 0$, then since $j_n(0) = 0$ for $n \geq 1$ (see (2.22)) we deduce that $\mu(\pi_1 U) = \epsilon' |j_0(0)|^2 = \epsilon'$ which is a stronger bound than required.

Lower Bound: By (2.30) we know that

$$j'_{n+1/2,1} - j'_{n-1/2,1} \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (2.39)$$

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 (2.39) we observe

$$|j'_{n(m)-1/2,1} - |2\pi\epsilon' \lambda \circ \rho(m)|| \leq 1/2 + \eta(m, \epsilon'), \quad (2.40)$$

where η is such that $\eta(m, \epsilon') \rightarrow 0$ as $m \rightarrow \infty$ for any fixed ϵ' . By using the same method as in (2.32) we find that

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

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

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

Combining this with (2.34) we see that, using (2.31),

$$\begin{aligned} |U_{m,n(m)}|^2 &= \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} |J_{n(m)-1/2}^2(2\pi\epsilon'\lambda \circ \rho(m))| \\ &\geq \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} \cdot (1 - \xi(m, \epsilon'))^2 \cdot |J_{n(m)-1/2,1}(j'_{n(m)-1/2,1})|^2 \\ &\geq \frac{n(m) - 1/2}{|\lambda \circ \rho(m)|} \cdot (1 - \xi(m, \epsilon'))^2 (\kappa(n(m) - 1/2)^{-1/3} + \mathcal{O}((n(m) - 1/2)^{-1}))^2. \end{aligned} \tag{2.41}$$

By (2.30), (2.40) and the fact that ρ is a standard ordering we know that (for ϵ' fixed)

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

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

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

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

Again what we have actually just proved (in more technical terms) is the following:

Theorem 2.3.5. *Let $U = [(B_f(\epsilon), \rho), (B_p, \tau)]$ where ρ is a frequency ordering of the Fourier basis. Then there is a constant $C_1 > 0$ such that for all $\epsilon \in (0, 1]$ 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 $\epsilon \in (0, 1]$ there exists an $M(\epsilon) \in \mathbb{N}$ such that for all $N \geq M$ we have the bound

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

2.4 The Impact of Incoherence on Sampling Strategies

We have shown that there is faster asymptotic incoherence for the Fourier-wavelet case ($\Theta(N^{-1})$) than for the Fourier-polynomial case ($\Theta(N^{-2/3})$). In this section we shall a quick example from [1] demonstrating how this difference is vital for choosing an effective sampling strategy.

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

$$f(x) = (1 - \cos(8\pi x)) \cdot \mathbb{1}_{[0,1]}(x), \quad x \in [-1, 1]. \quad (2.42)$$

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

$$\min_{x \in \ell^1(\mathbb{N})} \|x\|_1 \quad \text{subject to} \quad P_\Omega U x = P_\Omega \hat{f}. \quad (2.43)$$

We shall be using the SPGL1 package [12, 13] to solve (2.43) numerically. We focus on two choices of reconstruction bases⁶:

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

The coefficients of the decomposition of f into these two bases is shown in Figure 2.1. The coefficients in the polynomial expansion decay quickly, but there is little sparsity in the first 40 coefficients. On the other hand in the wavelet expansion there is large number of zeros in the first block of coefficients. This, combined with asymptotic incoherence, will enable us to subsample.

We shall be looking at two simple subsampling patterns and how they perform for each reconstruction basis. We shall be subsampling from the first 501 coefficients, and

⁶These bases were originally defined over $L^2[0, 1]$ but a simple change of variables is used to map $[0, 1]$ to $[-1, 1]$ to construct bases of $L^2[-1, 1]$.

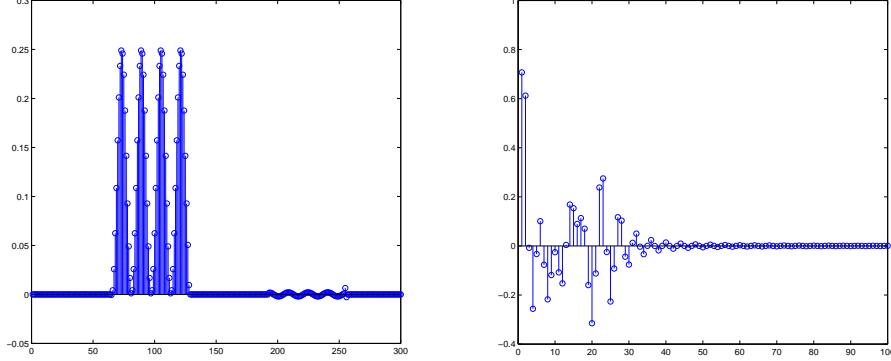
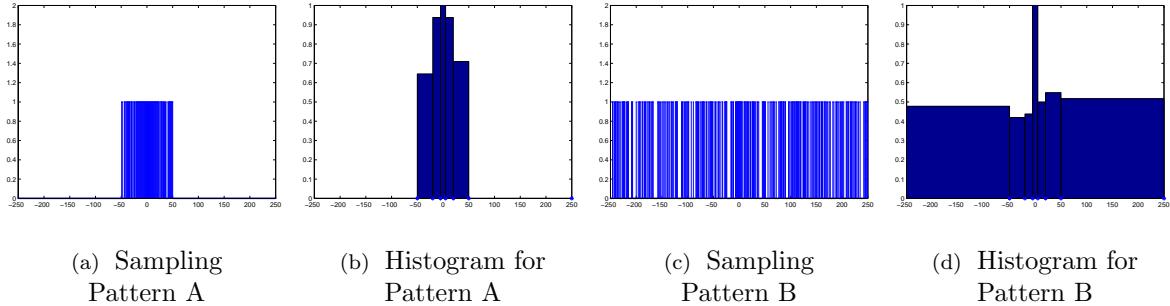
Figure 2.1: Coefficients of f when decomposed into different reconstruction bases.(a) First 300 coefficients (using a leveled ordering) of f in a Daubechies4 boundary wavelet expansion with $J=6$.(b) First 100 coefficients (using a natural ordering) of f in a Legendre polynomial expansion.

Figure 2.2: Two sampling patterns and their corresponding histograms.



since ρ is a frequency ordering this means that these coefficients correspond to

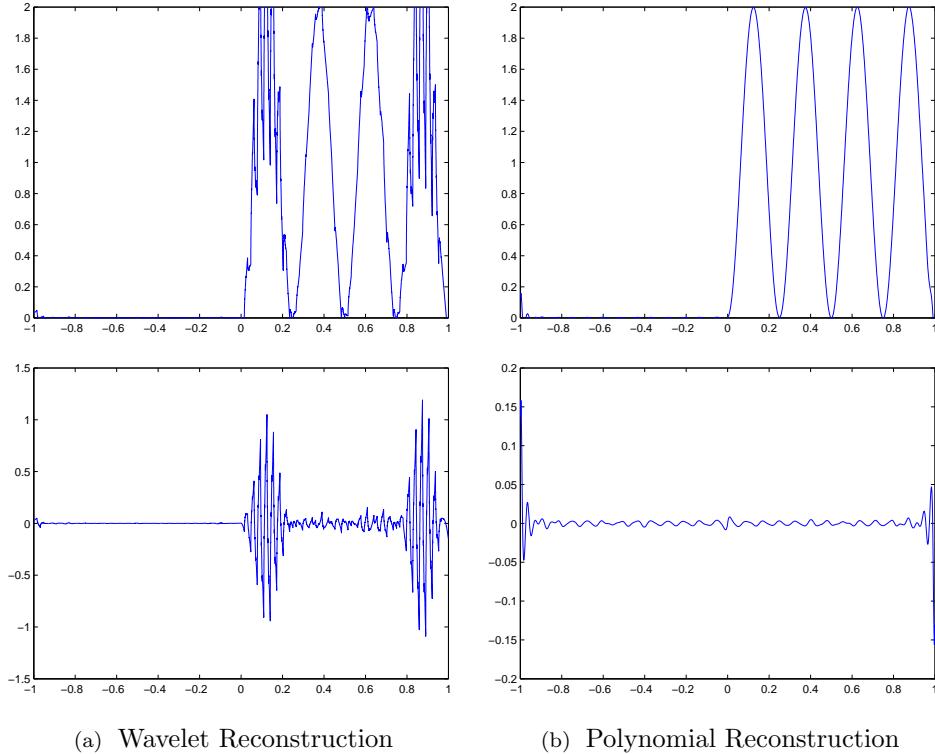
$$\{\lambda \circ \rho(m) : m = 1, \dots, 501\} = \{-250, -249, \dots, 249, 250\}.$$

If we were to sample all the 501 coefficients then we would achieve a highly accurate reconstruction from both bases⁷. We now consider two subsampling patterns, denoted as pattern A and pattern B which are presented in Figure 2.2, 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

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

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.

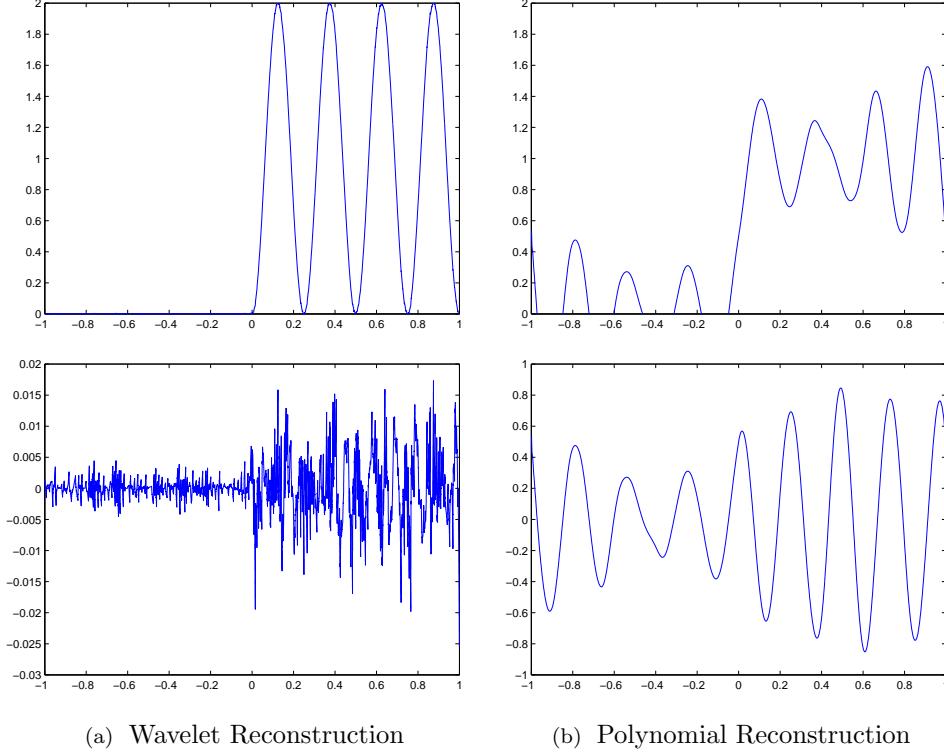
Figure 2.3: Reconstructions from Pattern A (above) with errors (below).



Let us first consider what happens when we use subsampling pattern A, which is shown in Figure 2.3. 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 2.4. 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 recon-

Figure 2.4: Reconstructions from Pattern B with errors.



struction, with an L^1 error of 7.29×10^{-1} . Due to poor sparsity and slow asymptotic incoherence, subsampling fails to be successful.

This again is another demonstration of the importance of the choice of reconstruction basis and that sampling methods based only on the density of Fourier samples will never reflect the true underlying picture.

2.5 Proving Optimality of the Coherence Bounds

In Sections 2.2 and 2.3 we derived bounds on the line coherences for the Fourier-wavelet case (say $U = U_w$) the Fourier polynomial case (say $U = U_p$)

$$\begin{aligned} \mu(\pi_N U_w), \mu(U_w \pi_N) &= \Theta(N^{-1}), & \mu(\pi_N U_p), \mu(U_p \pi_N) &= \Theta(N^{-2/3}), \\ \Rightarrow \mu(R_N U_w), \mu(U_w R_N) &= \Theta(N^{-1}), & \mu(R_N U_p), \mu(U_p R_N) &= \Theta(N^{-2/3}). \end{aligned} \tag{2.44}$$

This required that we work with a frequency ordering for the Fourier basis and leveled/natural orderings for the wavelet/polynomial basis. Our next goal is to show that no other orderings can improve upon the decay rates in (2.44).

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

Definition 2.5.1 (Relations on the set of orderings). *Let $U_1 := [(B_1, \rho_1), (B_2, \tau)]$, $U_2 := [(B_1, \rho_2), (B_2, \tau)]$. If*

$$\mu(R_N U_1) = \mathcal{O}(\mu(R_N U_2)), \quad N \rightarrow \infty, \quad (2.45)$$

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

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

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

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

Definition 2.5.2 (Optimal ordering). *Given the setup above, then any element of the optimal equivalence class is called an ‘optimal ordering of the basis pair (B_1, B_2) ’.*

It shall be shown in Lemma 2.7.2 that optimal orderings always exist . Notice that ρ is an optimal ordering if and only if for every other ordering ρ' we have $\rho \prec \rho'$. An optimal ordering has a corresponding optimal decay rate.

Definition 2.5.3 (Optimal decay rate). *Suppose $\rho : \mathbb{N} \rightarrow B_1$ is an optimal ordering for the basis pair (B_1, B_2) and $U = [(B_1, \rho), (B_2, \tau)]$. Then any decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ which satisfies $f(N) = \Theta(\mu(R_N U))$ is said to represent the ‘optimal decay rate’ of the basis pair (B_1, B_2) .*

At this point it is worth explaining why we define \prec in terms of block coherences and not line coherences. This is because $\mu(\pi_N U_1) = \mathcal{O}(\mu(\pi_N U_2))$ implies $\mu(R_N U_1) =$

$\mathcal{O}(\mu(R_N U_2))$ but not the other way around. Furthermore, $\mu(\pi_N U_1)$ is often not a decreasing function of N and a statement such as $\mu(\pi_N U_1) = \Theta(f(N))$ for a decreasing function f is not possible. However, the case when this does hold is very important to us:

Definition 2.5.4 (Optimal In lines). *Let $U = [(B_1, \rho), (B_2, \tau)]$. If $\mu(\pi_N U) = \Theta(f(N))$ for a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ then ρ is said to be ‘optimal in lines’ for the basis pair (B_1, B_2) .*

From (2.44) we see that the orderings used in these cases (frequency/leveled/natural) are optimal in lines by definition. We now show that the use of the word ‘optimal’ here is justified:

Proposition 2.5.5. *Let $U_1 = [(B_1, \rho_1), (B_2, \tau)]$ and $U_2 = [(B_1, \rho_2), (B_2, \tau)]$. If there exists a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ such that*

$$f(N) \leq \mu(\pi_N U_1), \quad N \in \mathbb{N}, \tag{2.46}$$

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

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

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

□

Corollary 2.5.6. *Let $U_1 = [(B_1, \rho_1), (B_2, \tau)]$ and $U_2 = [(B_1, \rho_2), (B_2, \tau)]$. If ρ_1 is optimal in lines , i.e. $\mu(\pi_N U_1) = \Theta(f(N))$ for a decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$, then $f(N) = \mathcal{O}(\mu(R_N U_2))$ and $\rho_1 \prec \rho_2$. Consequently, ρ_1 is optimal.*

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

Recall from Lemma 1.5.6 that $\mu(\pi_N U_1) = \Theta(f(N))$ implies $\mu(R_N U_1) = \Theta(f(N))$. Therefore $\mu(R_N U_1) = \mathcal{O}(\mu(R_N U_2))$ which by definition means $\rho_1 \prec \rho_2$.

□

We can now use Corollary 2.5.6 to immediately deduce the bounds in (2.44) are optimal for their respective basis pairs:

- Theorem 2.5.7.**
1. **Fourier-Wavelet Case:** Let $\epsilon \in I_{J,p}$. Frequency orderings are optimal for the basis pair $(B_f(\epsilon), B_w)$. Leveled orderings are optimal for the basis pair $(B_w, B_f(\epsilon))$. In both cases the optimal decay rate is $\Theta(N^{-1})$. These statements still hold with B_w replaced with the boundary wavelet basis B_{bw} and $\epsilon \in I_{J,p}$ by $\epsilon \in (0, 1]$. Likewise for the case of periodic wavelets and $\epsilon = 1$.
 2. **Fourier-Polynomial Case:** Let $\epsilon \in (0, 0.9]$. Frequency orderings are optimal for the basis pair $(B_f(\epsilon), B_p)$. Leveled orderings are optimal for the basis pair $(B_p, B_f(\epsilon))$. In both cases the optimal decay rate is $\Theta(N^{-2/3})$.

2.6 Theoretical Limits

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

Theorem 2.6.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then $\sum_N \mu(R_N U)$ diverges.*

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

$$\sum_{N=N'}^\infty |u_{N,j}|^2 \leq \sum_{N=N'}^\infty \mu(R_N U) \leq 1/4^2, \quad j \in \mathbb{N}. \quad (2.47)$$

Now define the vectors

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

Inequality (2.47) says that $\|v_j^2\|_2 \leq 1/4$ for every $j \in \mathbb{N}$. Since U is an isometry, we know its columns are normalised, i.e. $\|v_j\|_2 = 1$, and so we deduce $\|v_j^1\|_2 \geq 3/4$ for every $j \in \mathbb{N}$. Let $w_j := v_j^1 / \|v_j^1\|_2$, $j \in \mathbb{N}$. Since the $w_j \in \mathbb{C}^{N'-1}$ are all finite dimensional we claim that

$$\sup_{\substack{j, j' \in \{1, \dots, M\} \\ j \neq j'}} |\langle w_j, w_{j'} \rangle| \rightarrow 1, \quad \text{as } M \rightarrow \infty. \quad (2.48)$$

To see this, notice that for every $\epsilon > 0$, there exists a $\delta > 0$, such that for all $j \in \mathbb{N}$ the set $W_j(\epsilon) := \{w \in \mathbb{C}^{N'-1} : |\langle w_j, w \rangle| > 1 - \epsilon\}$ contains the open set $B_\delta(w_j)$ of radius

δ centered at w_j . It must be the case that there are $j_1, j_2 \in \mathbb{N}$, $j_1 \neq j_2$ such that $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \emptyset$, else the union

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

would form an open cover of the unit ball in $\mathbb{C}^{N'-1}$ with no finite subcover⁸, contradicting compactness of the unit ball in $\mathbb{C}^{N'-1}$. Since $B_{\delta/2}(w_{j_1}) \cap B_{\delta/2}(w_{j_2}) \neq \emptyset$, $w_{j_1} \in B_\delta(w_{j_2}) \subset W_{j_2}(\epsilon)$ and so $|\langle w_{j_1}, w_{j_2} \rangle| > 1 - \epsilon$. Since $\epsilon > 0$ was arbitrary we have proved (2.48).

Therefore, by (2.48) we know there exists $j_1, j_2 \in \mathbb{N}$, $j_1 \neq j_2$ such that $|\langle w_{j_1}, w_{j_2} \rangle| > 1/2$ and therefore we deduce that

$$|\langle v_{j_1}^1, v_{j_2}^1 \rangle| > \frac{1}{2} \|v_{j_1}^1\|_2 \|v_{j_2}^1\|_2 > \frac{3^2}{2 \cdot 4^2}. \quad (2.49)$$

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

$$|\langle v_{j_1}^2, v_{j_2}^2 \rangle| \leq \|v_{j_1}^2\|_2 \|v_{j_2}^2\|_2 \leq \frac{1}{4^2}. \quad (2.50)$$

Therefore, combining (2.49) with (2.50) gives us

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

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

Corollary 2.6.2. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then there does not exist an $\epsilon > 0$ such that*

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

Noting the above corollary and that $\mu(W) \geq N^{-1}$ is the best lower bound possible for any finite isometry $W \in \mathbb{C}^N \times \mathbb{C}^N$ it might be tempting to believe $\mu(R_N U) = \Theta(N^{-1})$ is the best decay rate we can achieve for an isometry $U \in \mathcal{B}(l^2(\mathbb{N}))$. However, it turns out that Theorem 2.6.1 cannot be improved without imposing additional conditions on U :

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

Theorem 2.6.3. Let $f, g : \mathbb{N} \rightarrow \mathbb{R}$ be any two strictly positive decreasing functions and suppose that $\sum_N f(N)$ diverges. Then there exists $U \in \mathcal{B}(l^2(\mathbb{N}))$ an isometry with

$$\mu(R_N U) \leq f(N), \quad \mu(UR_N) \leq g(N), \quad N \in \mathbb{N}. \quad (2.51)$$

Proof. The proof is constructive. We may assume without loss of generality that

$$f(N), g(N) \leq 1, \quad \forall N \in \mathbb{N}$$

We will construct a matrix $U = (u_{i,j})_{i,j \in \mathbb{N}}$ satisfying (2.51) with normalised columns, $v_j := (u_{i,j})_{i \in \mathbb{N}}, j \in \mathbb{N}$, all having disjoint support. With this in mind we partition \mathbb{N} as follows:

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

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

$$(v_j)_N = \begin{cases} (g(j)f(N))^{1/2}, & \text{if } \sum_{i=1}^{N-1} ((v_j)_i)^2 + g(j)f(N) \leq 1, \quad N \in \Omega_j, \\ (1 - \sum_{i=1}^{N-1} ((v_j)_i)^2)^{1/2}, & \text{if } \sum_{i=1}^{N-1} ((v_j)_i)^2 \leq 1, \\ \sum_{i=1}^{N-1} ((v_j)_i)^2 + g(j)f(N) \geq 1, & \quad N \in \Omega_j, \\ 0, & \text{Otherwise.} \end{cases} \quad (2.52)$$

It is immediate from the definition that v_j is supported on Ω_j and $((v_j)_N)^2 \leq f(N)g(j)$ for every $N, j \in \mathbb{N}$ which implies that (2.51) holds. Furthermore, it is easy to show by induction on N that $\|v_j\|_2 \leq 1$. Since f is decreasing and by the structure of the set Ω_j , $\sum_{N \in \Omega_j} f(N)$ diverges for every j and consequently there is an $N' \in \mathbb{N}$ such that

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

For $N \leq N'-1, N \in \Omega_j$ we fall into the first case of (2.52), however for $N = N'$ we fall into case 2, and therefore $\sum_{i=1}^{N'} ((v_j)_i)^2 = 1$. This means $\|v_j\|_2 = 1$ for every j and consequently U is an isometry. \square

Although this negative result shows that we cannot define an analogue of perfect incoherence for asymptotic incoherence, if we restrict our decay function to be a power

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

law, i.e. $f(N) := CN^{-\alpha}$ for some constants $\alpha, C > 0$ then the largest possible value of $\alpha > 0$ such that (2.51) holds for an isometry U is $\alpha = 1$, which is what we achieved in the Fourier-wavelet case.

2.7 Alternative Notions of Optimality

This is a good time to look further at why we work with Definition 2.5.1 as our notion of optimal decay in this thesis and discuss a possible alternative. One argument against the definition of optimality we use is that it is only unique up to constants, since it relies only on order notation. This is somewhat inconvenient if one wants to work with concrete estimates. Therefore it may be tempting to strengthen the notion of optimality in some way. One possible alternative would be the following:

Definition 2.7.1 (Best ordering). *Let (B_1, B_2) be a basis pair. Then any ordering $\rho : \mathbb{N} \rightarrow B_1$ is said to be a ‘best ordering’ if for any other ordering τ of B_2 and $U = [(B_1, \rho), (B_2, \tau)]$ we have that the function $g(N) := \mu(\pi_N U)$ is decreasing.*

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

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

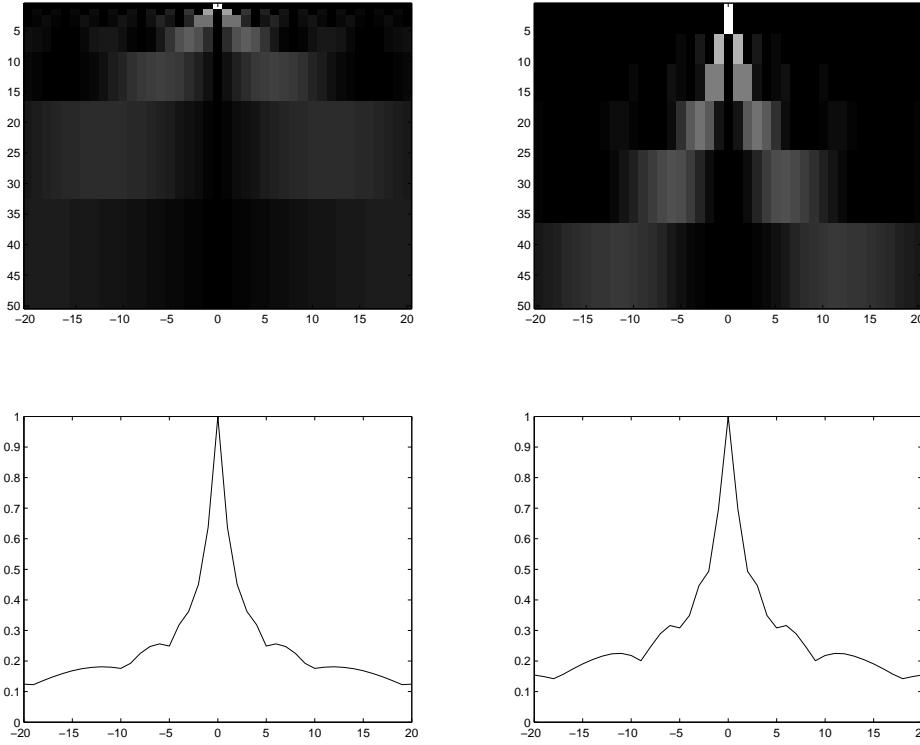
and we deduce that $\rho \prec \rho'$. This shows that any best ordering is optimal.

Lemma 2.7.2. *Suppose that we have a basis pair (B_1, B_2) . Then one of the following two results must hold:*

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

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

$$\sup_{N \in \mathbb{N} \setminus D} \mu(\pi_N U), \tag{2.53}$$



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

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

Figure 2.5: 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.

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

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

and for $N \geq 2$ we set

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

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 (2.53) is not attained for any $N \in \mathbb{N} \setminus D$. This means that if we remove finitely many elements from $\mathbb{N} \setminus D$ the supremum will remain unchanged. Therefore if N' is the largest natural number in D we find that

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

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

Lemma 2.7.3. *Suppose that we have a basis pair (B_1, B_2) with two orderings $\rho : \mathbb{N} \rightarrow B_1$, $\tau : \mathbb{N} \rightarrow B_2$ of B_1, B_2 respectively. If $U = [(B_1, \rho), (B_2, \tau)]$ satisfies*

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

then a best ordering exists.

Proof. The supremum (2.53) is always attained and therefore we fall into case 1 of the previous lemma. \square

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

Therefore, why did we not work with the definition of best ordering as our notion of optimality instead? The answer to this question is that best orderings are more exotic and far less simple to describe than optimal orderings. Figure 2.5 shows the Fourier-wavelet case where the best orderings are wavelet-dependent, even though we can describe optimal orderings in a wavelet-independent manner using frequency orderings. Although the difference between best orderings is very minor in Figure 2.5, this difference becomes considerable when working with the higher dimensional Fourier-wavelet cases.

2.8 Using Asymptotic Incoherence to Derive Standard Sampling Rates

Here we cover how the results in this chapter can be used to prove some sampling results. Recall that, thanks to [15] we already established a link between the block coherences and multilevel sampling using (1.28). In the Fourier-Wavelet case we have $\mu(R_N U), \mu(UR_N) = \Theta(N^{-1})$ and therefore we deduce

$$\begin{aligned} \mu_{\mathbf{N}, \mathbf{M}}(k, l) &\leq \sqrt{\min(\mu(R_{N_{k-1}+1}U), \mu(UR_{M_{l-1}+1})) \cdot \mu(R_{N_{k-1}+1}U)} \\ &\lesssim \frac{1}{\sqrt{N_{k-1}} \cdot \max(\sqrt{N_{k-1}}, \sqrt{M_{l-1}})}. \end{aligned} \quad (2.54)$$

To make this easier to handle we specify the sampling/sparsity regions that define N_k/M_{l-1} . Naturally we shall use the structure we have already been given and group the wavelet levels together:

$$M_1 = 1, \quad M_2 = 2, \quad M_3 = 4, \quad \dots, \quad M_l = 2^{l-1}, \quad \dots, \quad (2.55)$$

For the Fourier basis there is no such structure (only frequency size) and therefore we mimick the wavelet basis to keep the calculations clean:

$$N_1 = 1, \quad N_2 = 2, \quad N_3 = 4, \quad \dots, \quad N_k = 2^{k-1}, \quad \dots, \quad (2.56)$$

With this grouping (2.54) becomes:

$$\mu_{\mathbf{N}, \mathbf{M}}(k, l) \lesssim 2^{-k/2} \min(2^{-k/2}, 2^{-l/2}) = 2^{-k} \min(1, 2^{(k-l)/2}). \quad (2.57)$$

which, when applied to (1.25) gives the following bounds on multilevel sampling

$$m_k \gtrsim \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r'} \min(1, 2^{(k-l)/2}) \cdot s_l \right) \cdot \log(N), \quad k = 1, \dots, r. \quad (2.58)$$

If we remove the min term here we get the simple lower bound

$$\begin{aligned} m_k &\gtrsim \log(\epsilon^{-1}) \cdot s \cdot \log(N), \quad k = 1, \dots, r, \\ m = m_1 + \dots + m_r &\gtrsim sr \cdot \log(\epsilon^{-1}) \cdot \log(N). \end{aligned} \quad (2.59)$$

Since $M_r \leq N$ we see that $r \lesssim \log N$ and a weaker but simpler version of (2.58) is

$$m \gtrsim s \cdot \log(\epsilon^{-1}) \cdot \log^2(N) \quad (2.60)$$

Note that this is very similar to the standard $s \log(\epsilon^{-1}) \log(N)$ sampling rate of (1.7) expect it is out by a factor of $\log(N)$. This additional $\log(N)$ factor suggests that this approach of bounding the $\mu_{\mathbf{N}, \mathbf{M}}(k, l)$ is suboptimal.

The weak link in the argument above is the wasteful replacement of the $\min(1, 2^{(k-l)/2})$ term by 1. This originates from bounding the cross term

$$\mu(P_{N_{k-1}}^{N_k} U P_{M_{l-1}}^{M_l}),$$

by

$$\min(\mu(R_{N_{k-1}+1}U), \mu(UR_{M_{l-1}+1})) \lesssim \min(2^{-k}, 2^{-l}).$$

Therefore, this suggests that this bound is not optimal and can possibly be improved upon.

Work by Adcock, Hansen and Roman [31] demonstrated that, in the case of 1D Haar wavelets this term can be bounded by

$$\mu(P_{N_{k-1}}^{N_k} U P_{M_{l-1}}^{M_l}) \lesssim 2^{-k} 2^{-|k-l|}, \quad (2.61)$$

leading to the estimate

$$m_k \gtrsim \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r'} 2^{|k-l|/2} \cdot s_l \right) \cdot \log(N), \quad k = 1, \dots, r. \quad (2.62)$$

Summing this over k therefore gives a geometric sum for each s_l which can be bounded by a universal constant leaving us with the familiar bound

$$m = m_1 + \dots + m_r \gtrsim s \cdot \log(\epsilon^{-1}) \cdot \log(N), \quad k = 1, \dots, r. \quad (2.63)$$

Therefore the argument in [31] allows us to remove the $\log(N)$ factor and derive the classical $s \log(N)$ rate.

Here an extension of (2.61) is presented here so that it covers all one-dimensional Daubechies wavelet cases:

Lemma 2.8.1. *Let $U = [(B_f(\epsilon), \rho), (B_2, \tau)]$, where B_2 is one of the wavelet bases*

B_w, B_{pw}, B_{bw} , ρ is a frequency ordering and τ is a leveled ordering. Then there exists a constant $C > 0$, such that

$$|U_{NM}|^2 \leq C \cdot N^{-1} \min(MN^{-1}, NM^{-1}), \quad N, M \in \mathbb{N}. \quad (2.64)$$

Proof. The results $\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1})$ that has already been proved in this section shows that there exists a constant $C > 0$ such that

$$\begin{aligned} |U_{NM}|^2 &\leq \min(\mu(\pi_N U), \mu(\pi_M U)) \\ &\leq C \cdot \min(N^{-1}, M^{-1}) = C \cdot N^{-1} \cdot \min(1, NM^{-1}). \end{aligned} \quad (2.65)$$

Consequently, we need only to prove that, for a different constant $C' > 0$,

$$\mu(U_{NM}) \leq C' \cdot N^{-2} M, \quad N > M. \quad (2.66)$$

We use the explicit formula for U_{NM} we have used for the (periodic) wavelet case¹⁰:

$$|U_{NM}|^2 = |2^{-j(M)} \epsilon \mathcal{F}\psi(\epsilon 2^{-j(M)} \rho \circ \lambda(N))|^2. \quad (2.67)$$

To bound this term we use the stronger result from Lemma 2.2.3, namely

$$|\mathcal{F}\psi(\omega)| \leq \frac{K}{|\omega|}. \quad (2.68)$$

Applying this bound to (2.67) gives

$$|U_{NM}|^2 \leq 2^{j(M)} \epsilon^{-1} |\rho \circ \lambda(N)|^{-2} K^2.$$

Since τ is a leveled ordering we have that $2^{j(M)} = \Theta(M)$. Likewise since ρ is a frequency ordering we know that $|\rho \circ \lambda(N)| = \Theta(N)$. The result therefore follows. \square

Observe the crucial role that the Fourier transform decay (2.68) played in this result. Fourier transform decay (and hence wavelet smoothness) will have a greater role to play in the next chapter.

Despite these results showing that the Line/Block coherences do not give the full picture of coherence and the best sampling results, they get very close on their own, only missing out by a $\log(N)$ factor from the standard $s \log(N)$ result (2.60). More-

¹⁰We work with the ψ term only for simplicity.

over in cases where we can show optimality, they are a useful tool for identifying poor reconstruction bases, as in the polynomial case.

Chapter 3

Multidimensional Cases

In this section the coherence structures of two classes of multidimensional inverse problems are studied. Firstly, after some more theoretical tools are built up in Section 3.1, the general case of Tensor - Tensor basis pairs are considered in Section 3.2. Next the more specialised class of Fourier-Separable wavelets in considered in Section 3.3. Finally some CS numerics on the multidimensional cases are presented in Section 3.4.

3.1 Tools for Finding Optimal Orderings

Before we can tackle the multidimensional cases we will need to develop a few additional tools to simplify the analysis. The first is a formalisation of the technique used to prove the bounds on $\mu(\pi_N U), \mu(U\pi_N)$ for the one-dimensional Fourier-wavelet cases.

Definition 3.1.1. 1.) Suppose $F : S \rightarrow \mathbb{R}_{>0}$ satisfies $|\{x \in S : 1/F(x) \leq K\}| < \infty$ for all $K > 0$, $\sigma : \mathbb{N} \rightarrow S$ is consistent with $1/F$ and $F(\sigma(N)) \rightarrow 0$ as $N \rightarrow \infty$. Then any decreasing function $f : \mathbb{N} \rightarrow \mathbb{R}_{>0}$ such that $f(N) = \Theta(F \circ \sigma(N))$ is said to ‘represent the fastest decay of F ’.

2.) Suppose (B_1, B_2) is a basis pair and $\iota : S \rightarrow B_1$ a bijection. If there exists a function $F : S \rightarrow \mathbb{R}_{>0}$ and a constant $C_1 > 0$ such that

$$\sup_{g \in B_2} |\langle \iota(s), g \rangle|^2 \leq C_1 \cdot F(s), \quad \forall s \in S, \tag{3.1}$$

then F is said to ‘dominate the optimal decay of (B_1, B_2) ’. If the inequality is reversed we say F is ‘dominated by the optimal decay of (B_1, B_2) ’. Furthermore,

if F both dominates and is dominated by the optimal decay of (B_1, B_2) then F is said to ‘characterise the decay of (B_1, B_2) ’.

Example 3.1.2. Recall the formula (2.9) from the 1D Fourier-Wavelet case:

$$\mu(U\pi_N) = \epsilon \sup_{m \in \mathbb{Z}} 2^{-j(N)} |\mathcal{F}\psi(2^{-j(N)}\epsilon m)|^2. \quad (3.2)$$

This lead us to deduce that the decay of $\mu(\pi_N U)$ for $(B_{\text{pw}}, B_f(\epsilon))$ was essentially determined by the decay of $2^{-j(N)}$ where (for periodic wavelets) the function $j : \mathbb{N} \rightarrow \mathbb{R}$ satisfied

$$j(N) = \begin{cases} k, & 2^{k-1} < N \leq 2^k, \\ 0, & N = 1. \end{cases} \quad (3.3)$$

If, ignoring scaling function terms for simplicity, we rephrase (3.2) as

$$\sup_{m \in \mathbb{Z}} |\langle \psi_{j,k}, \chi_m \rangle|^2 = \epsilon 2^{-j} \sup_{m \in \mathbb{Z}} |\mathcal{F}\psi(2^{-j}\epsilon m)|^2, \quad (3.4)$$

then, with a little extra work bounding the supremum, one can show that the function $F(\psi_{j,k}) = 2^{-j}$ characterises the decay of $(B_{\text{pw}}, B_f(\epsilon))$ where $\iota : B_{\text{pw}} \rightarrow B_{\text{pw}}$ is the identity map.

Lemma 3.1.3. 1): Suppose $F : S \rightarrow \mathbb{R}_{>0}$ dominates the optimal decay of (B_1, B_2) , $\sigma : \mathbb{N} \rightarrow S$ is consistent with $1/F$ and $U = [(B_1, \iota \circ \sigma), (B_2, \tau)]$ with ι as in (3.1). Then if f represents the fastest decay of F then $\mu(\pi_N U) = \mathcal{O}(f(N))$.

2): If F is instead dominated by the optimal decay of (B_1, B_2) then $f(N) = \mathcal{O}(\mu(\pi_N U))$.

3): If F now characterizes the optimal decay of (B_1, B_2) then $\mu(\pi_N U) = \Theta(f(N))$ and ρ is a optimal in lines for the basis pair (B_1, B_2) if and only if $F(\iota^{-1} \circ \rho(N)) = \Theta(f(N))$.

Proof. 1.) From (3.1) we have ,

$$\sup_{g \in B_2} |\langle \iota \circ \sigma(N), g \rangle|^2 = \mu(\pi_N U) \leq C_1 \cdot F \circ \sigma(N), \quad \forall N \in \mathbb{N}.$$

Therefore $\mu(\pi_N U) = \mathcal{O}(F \circ \sigma(N)) = \mathcal{O}(g(N))$. The argument for 2.) follows the same lines.

3.) Results 1.) and 2.) combined gives $\mu(\pi_N U) = \Theta(f(N))$. Let $U' = [(B_1, \rho), (B_2, \tau)]$. Using corollary 2.5.6 we see that ρ is optimal in lines if and only if $\mu(\pi_N U') = \Theta(f(N))$.

Since F characterises the decay there exists $C_1, C_2 > 0$ such that

$$C_1 \cdot F(\iota^{-1} \circ \rho(N)) \leq \sup_{g \in B_2} |\langle \iota \circ \iota^{-1} \circ \rho(N), g \rangle|^2 = \mu(\pi_N U') \leq C_2 \cdot F(\iota^{-1} \circ \rho(N)), \quad (3.5)$$

therefore $\mu(\pi_N U) = \Theta(F(\iota^{-1} \circ \rho(N)))$. Therefore ρ is optimal in lines if and only if $\mu(\pi_N U) = \Theta(F(\iota^{-1} \circ \rho(N))) = \Theta(f(N))$. \square

Lemma 3.1.3 tells us that once we have found a function that characterises the optimal decay we can use it to construct optimal orderings and find the optimal decay rate. In the case of example 3.1.2 we found that (from the proof of Theorem 2.2.1) the (fastest) decay of $F(\psi_{j,k}) = 2^{-j}$ was represented by N^{-1} and therefore $\mu(\pi_N U) = N^{-1}$.

3.2 Multidimensional Tensor Cases

We recall the definition of a d -dimensional tensor basis B^d derived from a one dimensional basis B from Definition 1.7.5. In this section we aim to find optimal decay rates for multidimensional basis pairs of the form (B_1^d, B_2^d) where the analogous one dimensional case (B_1, B_2) is already understood.

3.2.1 General Estimates

Now suppose that we have two one-dimensional bases B_1, B_2 with corresponding optimal orderings ρ_1, ρ_2 . Let ρ_1^d, ρ_2^d be the d -dimensional indexings induced by ρ_1, ρ_2 of the bases B_1^d, B_2^d . What are optimal orderings of the basis pair (B_1^d, B_2^d) and what is the resulting optimal decay rate? Some insight is given by the following Lemma:

Lemma 3.2.1. *Let (B_1, B_2) be a pair of bases with corresponding tensor bases B_1^d, B_2^d . Let ρ_1 be an ordering of B_1 which is optimal in lines and ρ_1^d be the d -dimensional indexing induced by ρ_1 . Finally, for some ordering τ of B_2 , let $U = [(B_1, \rho_1), (B_2, \tau)]$. Then if f represents the optimal decay rate corresponding to the basis pair (B_1, B_2) we have, for some constants $C_1, C_2 > 0$,*

$$\prod_{i=1}^d C_1^d \cdot f(n_i) \leq \sup_{g \in B_2^d} |\langle \rho_1^d(n), g \rangle|^2 = \prod_{i=1}^d \mu(\pi_{n_i} U) \leq \prod_{i=1}^d C_2^d \cdot f(n_i), \quad n \in \mathbb{N}^d. \quad (3.6)$$

Consequently, if we let $\iota := \rho_1^d$ then $F(n) := \prod_{i=1}^d f(n_i)$ characterizes the optimal decay of (B_1^d, B_2^d) .

Proof. Let τ^d denote the d -dimensional indexing induced by τ . Then by breaking down the tensor product into terms and using the bijectivity of τ^d we have

$$\begin{aligned} \sup_{g \in B_2^d} |\langle \rho_1^d(n), g \rangle|^2 &= \sup_{m \in \mathbb{N}^d} |\langle \rho_1^d(n), \tau^d(m) \rangle|^2 = \sup_{m \in \mathbb{N}^d} \prod_{i=1}^d |\langle \rho_1(n_i), \tau(m_i) \rangle|^2 \\ &= \prod_{i=1}^d \sup_{m \in \mathbb{N}} |\langle \rho_1(n_i), \tau(m) \rangle|^2 = \prod_{i=1}^d \mu(\pi_{n_i} U). \end{aligned}$$

Therefore (3.6) follows from applying the definition of being optimal in lines to each term in the product. \square

Lemma 3.2.1 says that if we have an ordering that is optimal in lines for the basis pair (B_1, B_2) then we can use Lemma 3.1.3 to find all orderings that are optimal in lines for the corresponding tensor basis pair (B_1^d, B_2^d) . In particular, we have

Corollary 3.2.2. *We use the framework of the previous Lemma. Let $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ be consistent with $1/F$. Then an ordering ρ is optimal in lines for the basis pair (B_1^d, B_2^d) if and only if*

$$F((\rho_1^d)^{-1} \circ \rho(N)) = \Theta(F(\sigma(N))), \quad N \rightarrow \infty.$$

Suppose that we have an ordering ρ_1 of B_1 that is optimal in lines and has a decay rate that is a power of N , namely that $f(n) = n^{-\alpha}$ for some $\alpha > 0$. This is the case for the one dimensional examples we covered in Chapter 2. The above Lemma tells us that to find the optimal decay rate we should take an ordering $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ that is consistent with $1/F(n) := \prod_{i=1}^d 1/f(n_i) = \prod_{i=1}^d n_i^\alpha$ which is equivalent to being consistent with $1/F^{1/\alpha}(n) = \prod_{i=1}^d n_i$. This motivates the following:

Definition 3.2.3 (Corresponding to the hyperbolic cross). *Define $F_H : \mathbb{N}^d \rightarrow \mathbb{R}$ by $F_H(n) = \prod_{i=1}^d n_i$. Then we say a bijective function $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ ‘corresponds to the hyperbolic cross’ if it is consistent with F_H .*

The name ‘hyperbolic cross’ originates from its use in approximation theory [43, 44]. We now claim that if σ corresponds to the hyperbolic cross and $d \geq 2$, then

$$\prod_{i=1}^d \sigma(N)_i \sim \frac{(d-1)!N}{\log^{d-1}(N+1)} \quad \text{as } N \rightarrow \infty. \tag{3.7}$$

Next we proceed to prove this claim.

Definition 3.2.4. For $d \in \mathbb{N}$ let $f_d(x) = x \log^{d-1} x$ be defined on $[1, \infty)$. We define g_d as the inverse function of f_d on $[1, \infty)$, and so $g_d : [0, \infty) \rightarrow [1, \infty)$. Furthermore, we define

$$h_d(x) := \frac{x}{\log^{d-1}(x+1)}, \quad x \in [1, \infty). \quad (3.8)$$

Lemma 3.2.5. The following holds:

- 1.) $g_d(x)/h_d(x) \rightarrow 1$ as $x \rightarrow \infty$.
- 2.) Let $\tilde{f}(x) = x \log^{d-1} x + xp(\log(x)) + \beta$ with p a polynomial of degree at most $d-2$, $\beta \in \mathbb{R}$ and let \tilde{g} be its inverse function defined for large $x \in \mathbb{R}_+$. Then we also have $\tilde{g}(x)/h_d(x) \rightarrow 1$ as $x \rightarrow \infty$.

Proof. 1.) For notational convenience we shall prove the equivalent result

$$\frac{g_d(x) \log^{d-1}(x)}{x} \rightarrow 1 \quad \text{as } x \rightarrow \infty.$$

By taking logarithms we change the problem from studying the asymptotics of a fraction to the asymptotics of the difference

$$\log(g_d(x)) - \log(h_d(x)) = \log(g_d(x)) - \log x + (d-1) \log \log x \rightarrow 0 \quad \text{as } x \rightarrow \infty. \quad (3.9)$$

With this in mind we notice that the function $\log(g_d)$ (defined on $[0, \infty)$) is the inverse function of $e_d(x) := f_d(\exp(x)) = x^{d-1} \exp x$ (defined on $[0, \infty)$).

Notice that for x large we have $e_d(x - (d-1) \log x) = \frac{(x - (d-1) \log x)^{d-1}}{x^{d-1}} \exp(x) \leq \exp(x)$ which implies that $x - (d-1) \log x \leq \log(g_d(\exp(x)))$. Now if we let $\epsilon > 0$ then we deduce that

$$e_d(x - (d-1) \log x + \epsilon) = \frac{(x - (d-1) \log x + \epsilon)^{d-1}}{x^{d-1}} \exp(x + \epsilon) \geq \exp(x) \quad \text{for } x \text{ large.}$$

This implies that $x - (d-1) \log x + \epsilon \geq \log(g_d(\exp(x)))$ for x large. We therefore conclude that for all x sufficiently large we have

$$x - (d-1) \log x \leq \log(g_d(\exp(x))) \leq x - (d-1) \log x + \epsilon,$$

from which (3.9) follows since $\epsilon > 0$ is arbitrary.

- 2.) Notice that by part 1. it suffices to show that $\tilde{g}(x)/g_d(x) \rightarrow 1$ as $x \rightarrow \infty$.

Again, we shall show this by taking logarithms, reducing the proof to showing

$$\log(\tilde{g}(x)) - \log(g_d(x)) \rightarrow 0 \quad \text{as } x \rightarrow \infty.$$

Notice that $\log(\tilde{g}(x))$ is the inverse function, defined for large x , of

$$\tilde{e}(x) := \tilde{f}(\exp(x)) = x^{d-1} \exp(x) + p(x) \cdot \exp(x) + \beta,$$

Then since

$$\tilde{e}'(x) = x^{d-1} \exp(x) + ((d-1) \cdot x^{d-2} + p'(x) + p(x)) \cdot \exp(x),$$

we can use the hypothesis that p is of a lower order than x^{d-1} to show that for every $\epsilon > 0$, there is an $L(\epsilon) > 0$ such that for all $x \geq L(\epsilon)$ we have $\epsilon \cdot \tilde{e}'(x - \epsilon) \geq |\tilde{e}(x) - e_d(x)| = |p(x) \cdot \exp(x) + \beta|$. We therefore deduce from the mean value theorem that for $x \geq \exp(L(\epsilon))$ we have

$$\begin{aligned} \tilde{e}(\log(g_d(x)) - \epsilon) &\leq e_d(\log(g_d(x))) = x \leq \tilde{e}(\log(g_d(x)) + \epsilon) \\ &\Rightarrow \log(g_d(x)) - \epsilon \leq \log(\tilde{g}(x)) \leq \log(g_d(x)) + \epsilon, \end{aligned}$$

where we applied $\log(\tilde{g})$ to the inequality in the last step (this preserves the inequality since $\log(\tilde{g})$ is an increasing function of x for x large). \square

Lemma 3.2.6. 1). For every $d \in \mathbb{N}$ we have

$$R_N := \sum_{i=1}^N \frac{1}{i} (\log(N) - \log(i))^d = \frac{1}{d+1} \log^{d+1} N + \mathcal{O}(\log^d N), \quad N \rightarrow \infty. \quad (3.10)$$

2). Let $S_d(N)$ for $d, N \in \mathbb{N}$ be defined by

$$S_d(N) := \#\left\{m \in \mathbb{N}^d : \prod_{i=1}^d m_i \leq N\right\}. \quad (3.11)$$

Then for every $d \in \mathbb{N}$, there exists polynomials $\underline{p}_d, \bar{p}_d$ both of degree $d-1$ with identical leading coefficient $1/(d-1)!$ such that

$$N\underline{p}_d(\log(N)) \leq S_d(N) \leq N\bar{p}_d(\log(N)). \quad (3.12)$$

3). If we let $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ correspond to the hyperbolic cross then (3.7) holds.

Proof. 1). Let $I_N := \int_1^N \frac{1}{x}(\log(N) - \log(x))^d dx$. Since the integrand is a decreasing function of x (with N fixed) we find that by the Maclaurin integral test that $0 \leq R_N - I_N \leq \log^d(N)$. This means that showing (3.10) is equivalent to showing that

$$\int_1^N \frac{1}{x}(\log(N) - \log(x))^d dx = \frac{1}{d+1} \log^{d+1} N + \mathcal{O}(\log^d N).$$

Now, by expanding out the factors of the integrand and integrating (recall that the integral of $x^{-1} \log^k x$ is $\frac{1}{k+1} \cdot \log^{k+1} x$) the integral becomes

$$\log^{d+1}(N) \cdot \sum_{i=0}^d \frac{1}{i+1} \binom{d}{i} (-1)^i.$$

Since $\frac{1}{i+1} \binom{d}{i} = \frac{1}{d+1} \binom{d+1}{i+1}$ we see that the sum simplifies to $\frac{1}{d+1}$ and we are done.

2). We use induction on the dimension d . The case $d = 1$ is immediate since $\underline{p}_1(x) = \bar{p}_1(x) = 1$ satisfies inequality (3.12). Therefore suppose that inequality (3.12) holds for dimension $d = k$. We shall extend the result to $d = k + 1$ using the equality:

$$S_{k+1}(N) = \sum_{i=1}^N S_k\left(\left\lfloor \frac{N}{i} \right\rfloor\right). \quad (3.13)$$

This equality follows from rewriting the set defining S_{k+1} as the following disjoint union:

$$\left\{ m \in \mathbb{N}^{k+1} : \prod_{i=1}^{k+1} m_i \leq N \right\} = \coprod_{j=1}^N \left\{ m \in \mathbb{N}^{k+1} : m_{k+1} = j, \prod_{i=1}^k m_i \leq \left\lfloor \frac{N}{j} \right\rfloor \right\}.$$

Upper Bound: We may assume without loss of generality that \bar{p}_k has all coefficients positive. Therefore, by replacing $\left\lfloor \frac{N}{i} \right\rfloor$ with $\frac{N}{i}$ and using the upper bound in (3.12), we can upper bound equation (3.13) by

$$\sum_{i=1}^N \frac{N}{i} \cdot \bar{p}_k\left(\log\left(\frac{N}{i}\right)\right) \leq N \sum_{i=1}^N \frac{1}{i} \cdot \bar{p}_k(\log(N) - \log(i)).$$

We can then get the required upper bound by applying part 1) of the lemma to each

term in the polynomial; for example the highest order term becomes

$$\sum_{i=1}^N \frac{N}{i} \cdot \frac{1}{(k-1)!} (\log(N) - \log(i))^{k-1} \leq \frac{N}{k!} \log^k N + CN \log^{k-1} N, \quad \forall N \in \mathbb{N},$$

for some constant $C > 0$ sufficiently large. The other terms in \bar{p}_k are handled similarly.

Lower Bound: Notice that without loss of generality we can assume all the coefficients of \underline{p}_k apart from the leading coefficient are negative. Using the lower bound in (3.12), we can lower bound equation (3.13) by

$$\sum_{i=1}^N \left\lfloor \frac{N}{i} \right\rfloor \cdot \bar{p}_k \left(\log \left(\left\lfloor \frac{N}{i} \right\rfloor \right) \right).$$

This means we can tackle the $< k-1$ order terms in the same way as in the upper bound since we can replace $\left\lfloor \frac{N}{i} \right\rfloor$ with $\frac{N}{i}$ (recall we have assumed these terms are negative). Now we are left with bounding the highest order term:

$$\begin{aligned} & \sum_{i=1}^N \left\lfloor \frac{N}{i} \right\rfloor \frac{1}{(k-1)!} (\log \left(\left\lfloor \frac{N}{i} \right\rfloor \right))^k \\ &= \sum_{i=1}^N \left\lfloor \frac{N}{i} \right\rfloor \frac{1}{(k-1)!} \cdot \left[\log \left(\frac{N}{i} \right) - \left(\log \left(\frac{N}{i} \right) - \log \left(\left\lfloor \frac{N}{i} \right\rfloor \right) \right)^k \right]. \end{aligned} \tag{3.14}$$

Therefore expanding out the binomial term, setting the sign of all terms except the first to be negative, and noticing $\log \left(\frac{N}{i} \right) - \log \left(\left\lfloor \frac{N}{i} \right\rfloor \right) \leq 1$ for every i, N we get the lower bound

$$\sum_{i=1}^N \left\lfloor \frac{N}{i} \right\rfloor \frac{1}{(k-1)!} \log^k \left(\frac{N}{i} \right) - \sum_{i=1}^N \sum_{j=0}^{k-1} \left\lfloor \frac{N}{i} \right\rfloor \binom{k}{j} \frac{1}{(k-1)!} \log^j \left(\frac{N}{i} \right).$$

From here we can replace $\left\lfloor \frac{N}{i} \right\rfloor$ by $\frac{N}{i}$ for the right term, $\left\lfloor \frac{N}{i} \right\rfloor$ by $\frac{N}{i} - 1$ on the left term and use part 1) of the lemma again to prove the lower bound.

3.) From the second part of the lemma we know that for some degree $d-1$ polynomials $\underline{p}_d, \bar{p}_d$ with leading coefficient $1/(d-1)!$ we have $N\underline{p}_d(\log(N)) \leq S_d(N) \leq N\bar{p}_d(\log(N))$. Now notice that if $m \in \mathbb{N}$ then because of consistency we must have $S_d(F_H(\sigma(m))) - 1 \leq m$ since σ must first list all the terms n in \mathbb{N}^d with $F_H(n) \leq F_H(\sigma(m)) - 1$ before listing $\sigma(m)$. Likewise we must have $m \leq S_d(F_H(\sigma(m)))$ since the $S_d(F_H(\sigma(m)))$ terms with $F_H(n) \leq F_H(\sigma(m)), n \in \mathbb{N}^d$ must be listed by σ first,

including m , before any others. Consequently we deduce

$$(F_H(\sigma(m)) - 1)\underline{p}_d(\log(F_H(\sigma(m)) - 1)) \leq m \leq F_H(\sigma(m))\bar{p}_d(\log(F_H(\sigma(m)))) \quad (3.15)$$

We now treat both sides separately. Looking at the LHS we get the estimate $F_H(\sigma(m)) - 1 \leq \tilde{g}_d(m)$, where $\tilde{g}_d(m)$ is the inverse function (defined for large m) of

$$\tilde{f}_d(x) := \frac{1}{(d-1)!}x \log^{d-1}(x) + (\text{degree } d-2 \text{ poly})(\log(x)),$$

and so we may apply part 2. of Lemma 3.2.5 to deduce $F_H(\sigma(m)) \leq h_d((d-1)!m) \cdot (1 + \epsilon(m))$, where $\epsilon(m) \rightarrow 0$ as $m \rightarrow \infty$. The right hand side is handled similarly to get the same asymptotic lower bound on $F_H(\sigma(m))$, namely $F_H(\sigma(m)) \geq h_d((d-1)!m) \cdot (1 + \epsilon(m))$, where $\epsilon(m) \rightarrow 0$ as $m \rightarrow \infty$. Since $\frac{h_d((d-1)!x)}{(d-1)!h_d(x)} \rightarrow 1$ as $x \rightarrow \infty$ the proof is complete. \square

(3.7) allows us to determine the optimal decay rate for when the optimal one dimensional decay rate is a power of N .

Theorem 3.2.7. *Returning to the framework of Corollary 3.2.2, if $f(n) = n^{-\alpha}$ for $n \in \mathbb{N}$, $F(n) = \prod_{i=1}^d f(n)$ for $n \in \mathbb{N}^d$ and $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ corresponds to the hyperbolic cross then*

$$F(\sigma(N)) = \left(\prod_{i=1}^d \sigma(N)_i \right)^{-\alpha} \sim ((d-1)! \cdot h_d(N))^{-\alpha}, \quad N \rightarrow \infty. \quad (3.16)$$

Consequently $h_d^{-\alpha}$ is representative of the optimal decay rate for the basis pair (B_1^d, B_2^d) . Furthermore, an ordering ρ is optimal in lines for the basis pair (B_1^d, B_2^d) if and only if

$$\prod_{i=1}^d \left((\rho_1^d)^{-1} \circ \rho(N) \right)_i = \Theta(h_d(N)), \quad N \rightarrow \infty. \quad (3.17)$$

Proof. (3.16) follows immediately from (3.7). This implies that $F \circ \sigma \approx h_d^{-\alpha}$. The statement on the optimal decay rate then follows from the characterization result from Lemma 3.2.1 applied to Lemma 3.1.3. The statement on optimality follows from Corollary 3.2.2. \square

Definition 3.2.8. *Using the framework of Lemma 3.2.1, any ordering $\rho : \mathbb{N} \rightarrow B_1^d$ such that (3.17) holds is called a ‘hyperbolic’ ordering. with respect to ρ_1 . Notice that by*

(3.16) that if $\sigma : \mathbb{N} \rightarrow \mathbb{N}^d$ corresponds to the hyperbolic cross then $\rho_1^d \circ \sigma$ is hyperbolic with respect to ρ_1 .

We now apply Theorem 3.2.7 to the one-dimensional cases we have already covered:

3.2.2 Fourier-Wavelet Case

Theorem 3.2.9. *We use the setup of Lemma 3.2.1. Suppose $B_1 = B_f(\epsilon)$, $B_2 = B_w$ for some fixed $\epsilon \in I_{J,p}$, ρ_1 is a frequency ordering of B_1 and τ_1 is a leveled ordering of B_2 . Let $U_d = [(B_1^d, \rho), (B_2^d, \tau)]$ where ρ, τ is hyperbolic with respect to ρ_1, τ_1 respectively. Then we have*

$$\mu(\pi_N U_d), \mu(U_d \pi_N) = \Theta(N^{-1} \log^{d-1}(N+1)), \quad N \rightarrow \infty. \quad (3.18)$$

The above also holds if the basis B_w is replaced by B_{bw} (or B_{pw}) and the condition $\epsilon \in I_{J,p}$ by $\epsilon \in (0, 1]$ (or $\epsilon = 1$).

Proof. Inequality (3.18) follows from applying Theorems 2.2.1, 2.2.5 to Theorem 3.2.7. \square

3.2.3 Fourier-Polynomial Case

Theorem 3.2.10. *We use the setup of Lemma 3.2.1. Suppose $B_1 = B_f(\epsilon)$, $B_2 = B_p$ for some fixed $\epsilon \in (0, 0.9]$, ρ_1 is a frequency ordering of the Fourier basis and τ_1 is the natural ordering of the polynomial basis. Let $U_d = [(B_1^d, \rho), (B_2^d, \tau)]$ where ρ, τ is hyperbolic with respect to ρ_1, τ_1 respectively. Then we have*

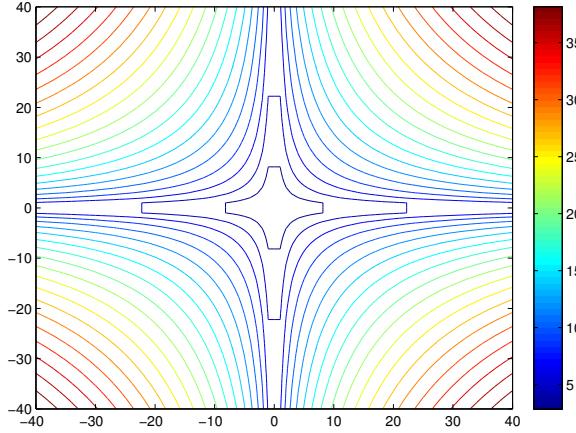
$$\mu(\pi_N U_d), \mu(U_d \pi_N) = \Theta(N^{-2/3} (\log^{d-1}(N+1))^{2/3}), \quad N \rightarrow \infty. \quad (3.19)$$

Proof. Inequality (3.19) follows from applying Theorems 2.3.2, 2.3.4 to Proposition 3.2.7. \square

3.2.4 Examples of Hyperbolic Orderings

The generalisation introduced by Definition 3.2.8, apart from allowing us to characterise all orderings that are optimal in lines, may seem to fulfil little other purpose. However, as we shall see in this section, this definition admits orderings which in specific cases

Figure 3.1: Hyperbolic Fourier Ordering in Two Dimensions:
A Contour Plot of H_2



are very natural and appear a little less abstract than an ordering derived from the hyperbolic cross.

Example 3.2.11. (*Hyperbolic Cross in \mathbb{Z}^d*) Our first example is unremarkable but nonetheless important. In d dimensions, take $B_1^d := B_f^d$ as a d -dimensional tensor Fourier basis. Recall we can identify this basis with \mathbb{Z}^d using the function λ_d . Suppose that we define a function $H_d : \mathbb{Z}^d \rightarrow \mathbb{R}$ by

$$H_d(m) = \prod_{i=1}^d |\max(|m_i|, 1)|, \quad (3.20)$$

and say that a bijective function $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ ‘corresponds to the hyperbolic cross in \mathbb{Z}^d ’ if it is consistent with H_d . Figure 3.1 shows the first few contour lines of H_d in two dimensions. With this definition we can then prove the analogous result of Lemma 3.2.6:

Lemma 3.2.12. Let $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ correspond to the hyperbolic cross and let h_d be as in (3.7). Then we have

$$\prod_{i=1}^d |\max(|\sigma(m)_i|, 1)| \sim \frac{(d-1)!}{2^d} \cdot h_d(m) \quad \text{as } m \rightarrow \infty. \quad (3.21)$$

Moreover, if ρ_1 is a standard ordering of B_f and $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ corresponds to the hyperbolic cross. Then $\lambda_d^{-1} \circ \sigma$ is a hyperbolic ordering with respect to ρ_1 .

Proof. Let $R_d(n)$ denote the number of lattice points in the hyperbolic cross of size n in

\mathbb{Z}^d , namely

$$R_d(n) := |\{m \in \mathbb{Z}^d : \prod_{i=1}^d \max(|m_i|, 1) \leq n\}|.$$

Call the set in the above definition $\mathcal{H}_d(n)$. If we remove the hyperplanes $\{m_i = 0\}$ for every i from $\mathcal{H}_d(n)$, we are left with 2^d quadrants in \mathbb{Z}^d which are congruent to set in equation (3.11). From the second part of Lemma 3.2.6 we therefore have

$$R_d(n) \geq 2^d n \underline{p}_d(\log(n)).$$

Next notice that the intersection of $\mathcal{H}_d(n)$ with each hyperplane $\{m_i = 0\}$ can be identified with $\mathcal{H}_{d-1}(n)$ and so we also have the upper bound

$$R_d(n) \leq 2^d n \bar{p}_d(\log(n)) + d \cdot R_{d-1}(n) \Rightarrow R_d(n) \leq n \bar{r}_d(\log(n)),$$

for some degree $d-1$ polynomial \bar{r}_d with leading coefficient $\frac{2^d}{(d-1)!}$. Combining the upper and lower bounds we see that for some polynomials $\underline{r}_d, \bar{r}_d$ of degree $d-1$ with leading coefficient $\frac{2^d}{(d-1)!}$ we have

$$n \underline{r}_d(\log(n)) \leq R_d(n) \leq n \bar{r}_d(\log(n)).$$

Therefore for $m \in \mathbb{N}$ since

$$R_d(H_d(\sigma(m)) - 1) \leq m \leq R_d(H_d(\sigma(m))),$$

we have

$$(H_d(\sigma(m)) - 1) \underline{r}_d(\log(H_d(\sigma(m)) - 1)) \leq m \leq H_d(\sigma(m)) \bar{r}_d(\log(H_d(\sigma(m)))).$$

Consequently we can apply Lemma 3.2.5 to both sides to derive (3.21) like in the proof of Lemma 3.2.6.

For the last part of the Lemma notice that since ρ_1 is a standard ordering then $\max(|\lambda_1 \circ \rho_1(N)|, 1) \approx N$. This means that the bounds on $\mu(\pi_N U)$ in Theorem 2.2.5 can be rephrased as

$$C_1 \cdot (\max(|n|, 1))^{-1} \leq \sup_{g \in B_w} |\langle \lambda_1^{-1}(n), g \rangle|^2 \leq C_2 \cdot (\max(|n|, 1))^{-1}, \quad n \in \mathbb{Z},$$

and by Lemma 3.2.1 this extends to the dD tensor case:

$$C_1^d \cdot \prod_{i=1}^d (\max(|n_i|, 1))^{-1} \leq \sup_{g \in B_w^d} |\langle \lambda_d^{-1}(n), g \rangle|^2 \leq C_2^d \cdot \prod_{i=1}^d (\max(|n_i|, 1))^{-1}, \quad n \in \mathbb{Z}^d. \quad (3.22)$$

This describes a characterization of the optimal decay of $(B_f^d(\epsilon), B_w^d)$. Lemma 3.1.3 tells us that $\lambda_d^{-1} \circ \sigma$ is optimal in lines for $(B_f^d(\epsilon), B_w^d)$, which by Theorem 3.2.7 is hyperbolic with respect to ρ_1 . \square

Example 3.2.13. (*Tensor Wavelet Ordering*) Now we look at an example of a less obvious hyperbolic ordering. We first introduce some notation to describe a tensor wavelet basis: For $j \in \mathbb{N}$, $k \in \mathbb{Z}$ let $\phi_{j,k}^0 := \phi_k$, $\phi_{j,k}^1 := \psi_{j,k}$. Now for $s \in \{0, 1\}^d$, $j \in \mathbb{N}^d$, $k \in \mathbb{Z}^d$ define

$$\Psi_{j,k}^s := \bigotimes_{i=1}^d \phi_{j_i, k_i}^{s_i}.$$

Then it follows that for $J \in \mathbb{N}$ fixed, we have

$$B_w^d := \left\{ \begin{array}{l} \text{Supp}(\phi_{j_i, k_i}^{s_i}) \cap (0, 1) \neq \emptyset \quad \forall i, \\ \Psi_{j,k}^s : s_i = 0 \Rightarrow j_i = J, \quad s_i = 1 \Rightarrow j_i \geq J \\ j \in \mathbb{N}^d, s \in \{0, 1\}^d, k \in \mathbb{Z}^d \end{array} \right\}, \quad (3.23)$$

The same approach can be applied to the boundary wavelet basis B_{bw} to generate a boundary tensor wavelet basis B_{bw}^d , although we must include the extra boundary terms, which can be done by letting $s \in \{0, 1, 2, 3\}^d$ where $\phi_{J,n}^2$ would be a boundary scaling function term and $\phi_{J,n}^3$ a boundary wavelet term.

Lemma 3.2.14. Let ρ_1 be any leveled ordering of a one-dimensional periodic wavelet basis B_{pw} . Setting $\bar{j} = \sum_{i=1}^d j_i$ define $F_{hyp} : B_{pw}^d \rightarrow \mathbb{R}$ by the formula

$$F_{hyp}(f) = \bar{j} \quad \text{if} \quad f = \Psi_{j,k}^s.$$

Then any ordering $\rho : \mathbb{N} \rightarrow B_{pw}^d$ that is consistent with F_{hyp} is a hyperbolic ordering with respect to ρ_1 .

Proof. We recall in that in the one-dimensional Fourier - (Periodic) Wavelet case, that for $\rho_1(N) = \phi_{j(N), k(N)}^{s(N)}$, $s \in \{0, 1\}$ we have

$$2^{j(N)} = \Theta(N), \quad N \rightarrow \infty.$$

Therefore, writing $\rho_1^d(m) = \Psi_{j(m), k(m)}^{s(m)}$,

$$C_1^d 2^{\overline{j(m)}} \leq \prod_{i=1}^d m_i \leq C_2^d 2^{\overline{j(m)}}, \quad m \in \mathbb{N}^d.$$

Consequently if we rewrite this with an actual ordering $\rho(N) = \Psi_{j(N), k(N)}^{s(N)}$ for $N \in \mathbb{N}$ we deduce

$$C_1^d 2^{\overline{j(N)}} \leq \prod_{i=1}^d ((\rho_1^d)^{-1} \circ \rho(N))_i \leq C_2^d 2^{\overline{j(N)}}, \quad (3.24)$$

and so we have reduced the problem to determining how $\overline{j(N)}$ scales with N . Notice that from our ordering of the wavelet basis that $\overline{j(N)}$ is a monotonically increasing function in N and moreover, for every value of $\overline{j(N)}$ there are $r_d(\overline{j(N)}) 2^{\overline{j(N)}}$ terms in B_w^d with this value of $\overline{j(N)}$ in the wavelet basis, where

$$r_d(N) := |\{(j, s) \in \mathbb{N}^d \times \{0, 1\}^d : \overline{j} = N, j_i(s_i - 1) = 0, \forall i = 1, \dots, d\}|.$$

This is where we have used that there are 2^j shifts of $\phi_{j,0}, \psi_{j,0}$ in B_{pw} . Notice that $r_d(N)$ is a polynomial of degree $d - 1$. With this in mind notice we can define, consistent for $n \in \mathbb{N}$, $n \geq J$,

$$T_d(x) = \sum_{i=J}^x r_d(i) 2^i := p_d(x) 2^x + \alpha_d,$$

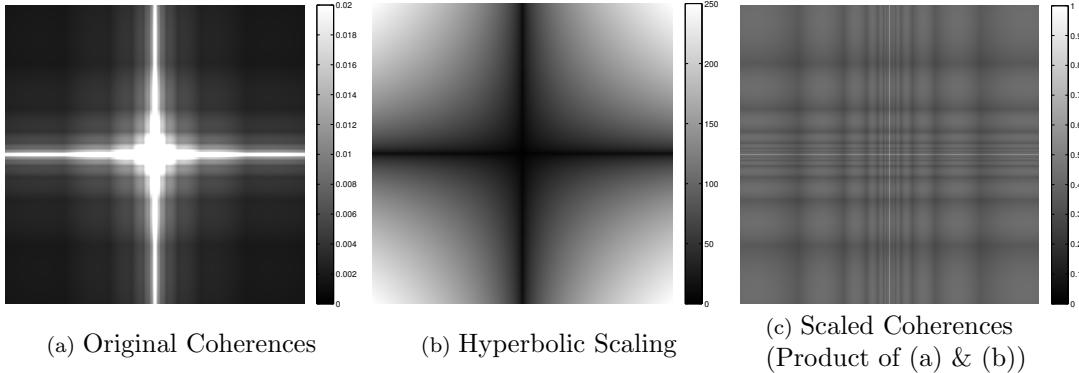
for some degree $d - 1$ polynomial p_d and constant α_d . This is possible by taking the formula for the geometric series expansion and differentiating repeatedly. By the consistency property of ρ we deduce the inequality

$$\begin{aligned} T_d(\overline{j(N)} - 1) \leq N \leq T_d(\overline{j(N)}) &\Rightarrow \overline{j(N)} - 1 \leq T_d^{-1}(N) \leq \overline{j(N)} \\ &\Rightarrow 2^{\overline{j(N)}-1} \leq 2^{T_d^{-1}(N)} \leq 2^{\overline{j(N)}}. \end{aligned}$$

Notice that $2^{T_d^{-1}(x)}$ is the inverse function of $T_d(\log_2 x)$ which is of the form $x \cdot p_d(\log_2 x) + \alpha$, Therefore, applying parts 2. & 3. of Lemma 3.2.5 gives, for some constants $D_1, D_2 > 0$ and N large,

$$(1 + \epsilon_1(N)) \cdot D_1 \cdot 2^{\overline{j(N)}} \leq g_d(N) \leq (1 + \epsilon_2(N)) \cdot D_2 \cdot 2^{\overline{j(N)}}, \quad (3.25)$$

Figure 3.2: 2D Tensor Fourier - Tensor Haar Incoherences



We show the subset $\{-250, -249, \dots, 249, 250\}^2 \subset \mathbb{Z}^2$. Notice that the scaled coherences have no vanishing values (no pure black) and no values that blow up (no pure white, barring the center value which = 1) indicating that we have characterised the coherence in terms of the hyperbolic scaling used. Formally this is shown by equation (3.22). The coherences shown in the Figure are square rooted to reduce contrast (i.e. we image $\sqrt{\mu(\pi_N U)}$ instead of $\mu(\pi_N U)$).

where $\epsilon_1(N), \epsilon_2(N) \rightarrow 0$ as $N \rightarrow \infty$. Combining this with (3.24) shows that we have a hyperbolic ordering. \square

3.2.5 Plotting Tensor Coherences

Let us consider a simple illustration of this theory applied to a 2D tensor Fourier-Wavelet case (B_f^2, B_w^2). We can identify the 2D Fourier Basis B_f^2 with \mathbb{Z}^2 using the function λ_2 (from (1.47)), so the row incoherences can also be identified with \mathbb{Z}^2 and therefore they can be imaged directly in 2D, as in Figure 3.2.

3.3 Multidimensional Fourier - Separable Wavelet Case

3.3.1 Defining the Separable Wavelet Basis

We repeat the notation of the one-dimensional case, with scaling function ϕ (in one dimension) & Daubechies wavelet ψ :

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

We can construct a d -dimensional scaling function Φ by taking the tensor product of ϕ with itself, namely

$$\Phi(x) := \left(\bigotimes_{j=1}^d \phi \right)(x) = \prod_{j=1}^d \phi(x_j), \quad x \in \mathbb{R}^d,$$

which has corresponding multiresolution analysis $(\tilde{V}_j)_{j \in \mathbb{Z}}$ with diagonal scaling matrix $A \in \mathbb{R}^{d \times d}$ with $A_{i,j} = 2\delta_{i,j}$.

Let $\phi^0 := \phi$, $\phi^1 := \psi$ and for $s \in \{0, 1\}^d$, $j \geq J$, $k \in \mathbb{Z}^d$ where $J \in \mathbb{N}$ is fixed define the functions

$$\Psi_{j,d}^s := \bigotimes_{i=1}^d \phi_{j,k_i}^{s_i}. \quad (3.26)$$

If we write (for $s \in \{0, 1\}^d \setminus \{0\}$, $j \geq J$)

$$W_j^s := \overline{\text{Span}\{\Psi_{j,k}^s : k \in \mathbb{Z}^d\}}.$$

Then it follows that

$$\tilde{V}_{j+1} = \tilde{V}_j \oplus \bigoplus_{s \in \{0,1\}^d \setminus \{0\}} W_j^s, \quad L^2(\mathbb{R}^d) = \overline{\tilde{V}_J \oplus \bigoplus_{\substack{s \in \{0,1\}^d \setminus \{0\} \\ j \geq J}} W_j^s}.$$

This corresponds to taking $2^d - 1$ wavelets for our basis in d dimensions (see [21]). As before we take the spanning functions from the above whose support has non-zero intersection with $[0, 1]^d$ as our basis B_2 (called a ‘separable wavelet basis’):

$$B_{\text{sep}}^d := \left\{ \begin{array}{ll} \text{Supp}(\phi_{j,k_i}^{s_i}) \cap (0, 1) \neq \emptyset & \forall i, \\ \Psi_{j,k}^s : & s = 0 \Rightarrow j = J, \\ & j \in \mathbb{N}, s \in \{0, 1\}^d, k \in \mathbb{Z}^d \end{array} \right\}, \quad (3.27)$$

Again, in the case $J = 0$, a periodic basis can be formed by periodising all elements in B_{sep}^d over $[0, 1]^d$, we denote this periodic separable wavelet basis by B_{psep}^d . We shall however focus on the standard separable wavelet basis B_{sep}^d in this Section.

Remark 3.3.1. *We can also construct a separable boundary wavelet basis in the same manner like in the one-dimensional case however, for the sake of simplicity, we stick to the above relatively simple construction throughout (although all the coherence results we cover here also hold for the separable boundary wavelet case as well).*

3.3.2 The Coherence Bounds

We recall a few key equalities from the one-dimensional case that will come in handy:

$$\mathcal{F}\phi_{j,k}(\omega) = e^{-2\pi i 2^{-j}k\omega} 2^{-j/2} \mathcal{F}\phi(2^{-j}\omega), \quad \mathcal{F}\psi_{j,k}(\omega) = e^{-2\pi i 2^{-j}k\omega} 2^{-j/2} \mathcal{F}\psi(2^{-j}\omega), \quad (3.28)$$

where \mathcal{F} here denotes the Fourier Transform, i.e. for $f \in L^2(\mathbb{R}^d)$ we define

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

Recall χ_k from Definition 1.30. We observe that by (3.26)

$$\begin{aligned} \langle \Psi_{j,k}^s, \chi_n \rangle &= \epsilon^{d/2} \cdot \mathcal{F}\Psi_{j,k}^s(\epsilon n) = \epsilon^{d/2} \prod_{i=1}^d \mathcal{F}\phi_{j,k_i}^{s_i}(\epsilon n_i), \quad n \in \mathbb{Z}^d, \\ &\Rightarrow \sup_{n \in \mathbb{N}^d} |\langle \Psi_{j,k}^s, \chi_n \rangle|^2 = \epsilon^d 2^{-dj} \cdot \prod_{i=1}^d \sup_{n \in \mathbb{N}} |\mathcal{F}\phi^{s_i}(\epsilon 2^{-j} n)|^2. \end{aligned} \quad (3.29)$$

By careful treatment of the product term we can determine the optimal decay of $(B_{\text{sep}}^d, B_{\text{f}}^d(\epsilon))$, using the following result:

Proposition 3.3.2. *There are constants $C_1, C_2 > 0$ such that for all $\epsilon \in I_{J,p}$, $\Psi_{j,k}^s \in B_{\text{sep}}^d$ we have*

$$C_1 \cdot \epsilon^d 2^{-dj} \leq \sup_{n \in \mathbb{N}^d} |\langle \Psi_{j,k}^s, \chi_n \rangle|^2 \leq C_2 \cdot \epsilon^d 2^{-dj}.$$

Consequently, fixing ϵ , the function $F_{\text{power}} : B_{\text{sep}}^d \rightarrow \mathbb{R}$ defined by $F_{\text{power}}(\Psi_{j,k}^s) = 2^{-dj}$ characterizes the optimal decay of $(B_{\text{sep}}^d, B_{\text{f}}^d(\epsilon))$.

Proof. Let $A = \max(\sup_{\omega \in \mathbb{R}} |\mathcal{F}\phi(\omega)|^2, \sup_{\omega \in \mathbb{R}} |\mathcal{F}\psi(\omega)|^2)$. Then (3.29) gives us the upper bound

$$\sup_{n \in \mathbb{N}^d} |\langle \Psi_{j,k}^s, \chi_n \rangle|^2 \leq \epsilon^d 2^{-dj} \cdot A^d.$$

This leaves the lower bound. This can be achieved if we can show that there exists constants $D_1, D_2 > 0$ such that for all $\epsilon \in I_{J,p}$ ¹

$$\mathcal{F}_1(\epsilon) := \sup_{n \in \mathbb{N}} |\mathcal{F}\phi(\epsilon 2^{-J} n)| \geq D_1, \quad \mathcal{F}_2(\epsilon) := \sup_{n \in \mathbb{N}} |\mathcal{F}\psi(\epsilon 2^{-J} n)| \geq D_2. \quad (3.30)$$

¹Notice that replacing J with $j \geq J$ below would have been redundant.

By the Riemann-Lebesgue Lemma the functions $\mathcal{F}_1, \mathcal{F}_2$ are continuous on $I_{J,p}$ and

$$\mathcal{F}_1(\epsilon) \rightarrow \sup_{\omega \in \mathbb{R}} |\mathcal{F}\phi(\omega)| > 0 \quad \text{as } \epsilon \rightarrow 0.$$

Likewise for \mathcal{F}_2 . Therefore $\mathcal{F}_1, \mathcal{F}_2$ can be extended to continuous functions over the closed interval $I_{J,p} \cup \{0\}$. Finally we notice that $\mathcal{F}_1(\epsilon) > 0, \mathcal{F}_2(\epsilon) > 0$ for every $\epsilon \in I_{J,p}$ otherwise we would deduce that ϕ or ψ has no support in $[0, 1]$ since the span of $B_f^d(\epsilon)$ covers $L^2[0, 1]$. This means that the infimums over $I_{J,p} \cup \{0\}$ are attained and are strictly positive, proving (3.30) and the lower bound. \square

Let $F_{\text{level}} : B_{\text{sep}}^d \rightarrow \mathbb{R}$ be defined by $F_{\text{level}}(\Psi_{j,k}^s) = j$. Lemma 3.1.3 tells that an ordering that is consistent with $1/F_{\text{power}}$, i.e. consistent with F_{level} will be optimal in lines.

Definition 3.3.3. *We say that an ordering $\rho : \mathbb{N} \rightarrow B_{\text{sep}}^d$ is ‘leveled’ if it is consistent with F_{level} .*

Lemma 3.3.4. *Let $\rho : \mathbb{N} \rightarrow B_{\text{sep}}^d$ or $\rho : \mathbb{N} \rightarrow B_{\text{psep}}^d$ be leveled. Then*

$$2^{dF_{\text{level}}(\rho(N))} = \Theta(N), \quad N \rightarrow \infty. \quad (3.31)$$

For B_{sep}^d . Let $a \in \mathbb{N}$ denote the length of the support of ϕ, ψ . Notice that for each $j \in \mathbb{N}$ and $s \in \{0, 1\}^d$, there are $(2^{j+1} + a - 1)^d$ shifts of $\Psi_{j,0}^s$ whose support lies in $[0, 1]^d$. For convenience we use the notation $j(N) := F_{\text{level}}(\rho(N))$ and shall also be using 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 $f \in B_{\text{sep}}^d, F_{\text{level}}(f) = j(N) - 1$ come before N in the leveled ordering and there are at least $(2^d - 1) \cdot 2^{dj(N)}$ of these terms, implying that

$$(2^d - 1) \cdot 2^{dj(N)} \leq N.$$

This completes the upper bound for $j(N) > J$. Likewise for every $N \in \mathbb{N}$ with $j(N) \geq J$ there can be no more than

$$2^d \cdot \sum_{i=J}^{j(N)} 2^{d(i+a)} \leq 2^d \cdot 2^{d(j(N)+a+1)} = 2^{d(a+2)} \cdot 2^{dj(N)},$$

terms such that $F_{\text{level}}(f) \leq j(N)$. This shows that $N \leq 2^{d(a+2)} \cdot 2^{dj(N)}$, completing the upper bound for $j(N) > J$. Extending (3.31) to all $N \in \mathbb{N}$ (i.e. $j(N) \geq J$) is trivial

since we have only omitted finitely many terms so a change of constants will suffice. \square

Theorem 3.3.5. *Any ordering ρ of B_{sep}^d that is leveled is optimal in lines for the basis pair $(B_{\text{sep}}^d, B_{\text{f}}^d(\epsilon))$. Furthermore, the optimal decay rate of $(B_{\text{sep}}^d, B_{\text{f}}^d(\epsilon))$ is represented by the function $f(N) = N^{-1}$.*

Proof. Lemma 3.1.3 applied to Proposition 3.3.2 tells us that ρ is optimal in lines and moreover the optimal decay rate is represented by $F_{\text{power}}(\rho(N))$ which by Lemma 3.3.4 is of order N^{-1} . \square

3.3.3 Ordering the Fourier Basis

We now want to find the optimal decay rate of $(B_{\text{f}}^d(\epsilon), B_{\text{sep}}^d)$ which means looking at orderings of the Fourier basis. It might be tempting to try and extend the definition of a frequency ordering from the one dimensional case. Recall as well that, using the function λ_d defined in (1.47), ordering $B_{\text{f}}^d(\epsilon)$ is equivalent to ordering \mathbb{Z}^d . Frequency orderings were motivated by ordering \mathbb{Z} by size. What size means in \mathbb{Z}^d for $d \geq 2$ is less clear. Furthermore, it turns out that the family of optimal orderings become wavelet dependent when $d \geq 3$. Herein lies the difficulty of covering the separable Fourier-wavelet case in general.

If we let $s \in \{0, 1\}^d$, $j \in \mathbb{N}$, $k \in \mathbb{Z}^d$, then in order to bound the coherence $\mu(\pi_N U)$ we need to be bounding terms of the form

$$|\langle \Psi_{j,k}^s, \lambda_d^{-1}(n) \rangle|^2 = \epsilon^d 2^{-dj} \prod_{i=1}^d |\mathcal{F}\phi^{s_i}(2^{-j}\epsilon n_i)|^2. \quad (3.32)$$

In the one-dimensional case in Chapter 2 the following decay property of the Fourier transform of the scaling function ϕ was used:

$$|\mathcal{F}\phi(\omega)|, |\mathcal{F}\psi(\omega)| \leq \frac{K}{|\omega|^{1/2}}, \quad (3.33)$$

Therefore let us first consider the case where we use (2.14) to bound every term in the product, giving us ($n \in \mathbb{Z}^d, n_i \neq 0, i = 1, \dots, d$)

$$|\langle \Psi_{j,k}^s, \lambda_d^{-1}(n) \rangle|^2 \leq \epsilon^d 2^{-dj} \prod_{i=1}^d \frac{K^2}{|\epsilon 2^j n_i|} = \frac{K^{2d}}{\prod_{i=1}^d |n_i|}, \quad (3.34)$$

making adjustments to prevent dividing by zero by using

$$\sup_{\omega \in \mathbb{R}} \max(|\mathcal{F}\phi(\omega)|, |\mathcal{F}\psi(\omega)|) \leq 1, \quad (3.35)$$

(for ϕ this follows from Proposition 1.11 in [22]. We extend this to ψ using equation (2.17)). This can then be rephrased as

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 \leq \frac{\max(K^{2d}, 1)}{\prod_{i=1}^d \max(|n_i|, 1)}, \quad n \in \mathbb{Z}^d. \quad (3.36)$$

This tells us that the function $F_{\text{hyp}} : \mathbb{Z}^d \rightarrow \mathbb{R}$, $F_{\text{hyp}}(n) = (\prod_{i=1}^d \max(|n_i|, 1))^{-1}$ dominates the optimal decay of $(B_f^d(\epsilon), B_{\text{sep}}^d)$ (see Definition 3.1.1 for the definition of domination). Therefore if we want to maximise the utility of this bound then we should use an ordering σ of \mathbb{Z}^d so that $\prod_{i=1}^d \max(|\sigma(N)_i|, 1)$ is increasing, namely an ordering corresponding to the hyperbolic cross in \mathbb{Z}^d (see Example 3.2.11). However, using such an ordering will not give us the N^{-1} decay rate that we got from the one dimensional case:

Proposition 3.3.6. *Let $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ correspond to the hyperbolic cross in \mathbb{Z}^d and define an ordering ρ of $B_f^d(\epsilon)$ by $\rho := \lambda_d^{-1} \circ \sigma$, where $\epsilon \in I_{J,p}$. Next let $U = [(B_f^d(\epsilon), \rho), (B_{\text{sep}}^d, \tau)]$ for any ordering τ and fix ϵ . Then*

$$\mu(R_N U) = \Theta(N^{-1} \log^{d-1} N), \quad N \rightarrow \infty.$$

As this result is primarily for motivation, its proof is left to the appendix.

Since this approach gives us suboptimal results, we return to our bound of (3.32). Instead of using (2.14) on every term in the product, why not just use it once on the term that give us the best decay instead? To bound the remaining terms we can simply use (3.35). This approach gives us the following bound

$$|\langle \Psi_{j,k}^s, \lambda_d^{-1}(n) \rangle|^2 \leq \epsilon^{d-1} 2^{-(d-1)j} \cdot \frac{K^2}{\max_{i=1,\dots,d} |n_i|}, \quad n \in \mathbb{Z}^d. \quad (3.37)$$

However it turns out than we can do better here; if we instead replace (2.14) by the stronger condition

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{d/2}}, \quad \omega \in \mathbb{R} \setminus \{0\}. \quad (3.38)$$

then we can obtain the following upper bound²

$$|\langle \Psi_{j,k}^s, \rho(N) \rangle|^2 \leq \epsilon^d 2^{-dj} \cdot \min_{i=1,\dots,d} \frac{K^{2d}}{|\epsilon 2^j n_i|^d} = \frac{K^{2d}}{\max_{i=1,\dots,d} |n_i|^d}. \quad (3.39)$$

Let us write $\|n\|_\infty := \max_{i=1,\dots,d} |n_i|$. The above can be rephrased as

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 \leq \frac{\max(K^{2d}, 1)}{\max(\|n\|_\infty^d, 1)}, \quad n \in \mathbb{Z}^d. \quad (3.40)$$

Therefore we deduce that $F_{\text{lin}} : \mathbb{Z}^d \rightarrow \mathbb{R}$, $F_{\text{lin}}(n) = (\max(\|n\|_\infty^d, 1))^{-1}$ dominates the optimal decay of $(B_f^d(\epsilon), B_{\text{sep}}^d)$. In fact it can be shown that F_{lin} also *characterizes* the optimal decay (i.e. a lower bound of the same form is possible) by using Lemma 2.2.4:

Proposition 3.3.7. *We fix the choice of wavelet basis B_{sep}^d and recall the function $\lambda_d : B_f^d(\epsilon) \rightarrow \mathbb{Z}^d$ from (1.47).*

1.) *Then there are constants $C_1(\phi) > 0, D(J) > 0$ such that for all $\epsilon \in I_{J,p}$ and $n \in \mathbb{Z}^d$ with $\|n\|_\infty \geq D\epsilon^{-1}$ we have*

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 \geq \frac{C_1}{\|n\|_\infty^d}. \quad (3.41)$$

Therefore (by fixing ϵ) the function F_{lin} is dominated by the optimal decay of $(B_f^d(\epsilon), B_{\text{sep}}^d)$.

2.) *Suppose that ϕ satisfies (3.38). Then there is a constant $C_2(\phi) > 0$ such that for all $\epsilon \in I_{J,p}$ and $n \in \mathbb{Z}^d$,*

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 \leq \frac{C_2}{\max(\|n\|_\infty^d, 1)}. \quad (3.42)$$

Therefore (by fixing ϵ) the function F_{lin} characterizes the optimal decay of $(B_f^d(\epsilon), B_{\text{sep}}^d)$.

Proof. 2.) Follows from (3.40).

1.) If we set $j = \lceil \log_2 \epsilon \|n\|_\infty \rceil + q$ for some $q \in \mathbb{N}$ fixed we observe that $|\epsilon 2^{-j} n_i| \in [0, 2^{-q}]$ for every $i = 1, \dots, d$ and, since we are using the max norm, $|\epsilon 2^{-j} n_i| \in [2^{-q-1}, 2^{-q}]$ for at least one i , say i' . Set $s_i = 0$ for $i \neq i'$ and $s_{i'} = 1$. Then, assuming $j \geq J$, by

²noting that (3.38) also holds for ψ by (2.17).

(3.32) we have the lower bound.

$$\begin{aligned} |\langle \Psi_{j,0}^s, \lambda_d^{-1}(n) \rangle|^2 &\geq \frac{2^{-d(q+1)}}{\|n\|_\infty^d} \prod_{i=1}^d |\mathcal{F}\phi^{s_i}(\epsilon 2^{-j} n_i)|^2 \\ &\geq \frac{2^{-d(q+1)}}{\|n\|_\infty^d} \cdot \inf_{\omega \in (2^{-q-1}, 2^{-q}]} |\mathcal{F}\psi(\omega)|^2 \cdot \inf_{\omega \in [0, 2^{-q}]} |\mathcal{F}\phi(\omega)|^{2(d-1)}. \end{aligned} \quad (3.43)$$

Recall that by Lemma 2.2.4 there exists a $q \in \mathbb{N}$ such that $L_q > 0$ and $\inf_{\omega \in [0, 2^{-q}]} |\mathcal{F}\phi(\omega)| > 0$ ³ and therefore (3.42) follows as long as $j \geq J$.

To ensure that $j = \lceil \log_2(\epsilon \|n\|_\infty) \rceil + q$ satisfies $j \geq J$ we must therefore impose the constraint that n is sufficiently large. $j \geq J$ is satisfied if

$$J \leq \log_2(\epsilon \|n\|_\infty) \Rightarrow \|n\|_\infty \geq 2^J \epsilon^{-1}.$$

□

Remark 3.3.8. If $d = 2$ then (3.38) always holds by Lemma 2.2.3. This means we have characterized every 2D Separable wavelet case (for Daubechies Wavelets).

Let $F_{\text{norm}}(n) := \max(\|n\|_\infty, 1)$. By Lemma 3.1.3 we know that if (3.38) holds then the optimal decay of $(B_f^d(\epsilon), B_{\text{sep}}^d)$ is determined by the fastest growth of F_{norm} . This motivates the following:

Lemma 3.3.9. Let $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ be consistent with F_{norm} . Then

$$\max(\|\sigma(N)\|_\infty, 1) = \Theta(N^{1/d}), \quad N \rightarrow \infty. \quad (3.44)$$

Proof. If $\|\sigma(N)\|_\infty = L \geq 2$, then σ must have enumerated beforehand all points m in \mathbb{Z}^d with $\|m\|_\infty \leq L - 1$ and there are $(2L - 1)^d$ of such points. This means that

$$N \geq (2L - 1)^d \Rightarrow \|\sigma(N)\|_\infty \leq \frac{N^{1/d} + 1}{2}, \quad N \in \mathbb{N}.$$

which proves the upper bound when $\|\sigma(N)\|_\infty = L \geq 2$. The lower bound is tackled similarly by noting σ must first list all $m \in \mathbb{Z}^d$ with $\|m\|_\infty \leq L$, including $\sigma(N)$ which shows

$$N \leq (2L + 1)^d \Rightarrow \|\sigma(N)\|_\infty \geq \frac{N^{1/d} - 1}{2}, \quad N \in \mathbb{N}.$$

□

³We are using the fact that $|\mathcal{F}\phi(0)| = 1$ and continuity of $\mathcal{F}\phi$ here which follows from $\phi \in L^1(\mathbb{R})$.

Definition 3.3.10 (Linear Ordering). *Any ordering $\rho : \mathbb{N} \rightarrow B_f^d(\epsilon)$ such that $\sigma = \lambda_d \circ \rho$ satisfies (3.44) is called a ‘linear ordering’.*

Theorem 3.3.11. *Assuming (3.38) holds for the scaling function corresponding to B_{sep}^d , an ordering ρ of $B_f^d(\epsilon)$ is optimal in lines for the basis pair $(B_f^d(\epsilon), B_{sep}^d)$ if and only if it is linear. Furthermore, the optimal decay rate of $(B_f^d(\epsilon), B_{sep}^d)$ is represented by the function $f(N) = N^{-1}$.*

Proof. If we apply part 2.) of Proposition 3.3.7 to Lemma 3.1.3 we know that if $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is consistent with $1/F_{lin} = F_{norm}^d$, i.e. consistent with F_{norm} , then $F_{lin}(\sigma(\cdot)) = 1/F_{norm}^d(\sigma(\cdot))$ represents the optimal decay rate. Lemma 3.3.9 tells us that this optimal decay is $1/(N^{1/d})^d = 1/N$. Furthermore, Lemma 3.1.3 says that an ordering ρ is optimal in lines for $(B_f^d(\epsilon), B_{sep}^d)$ if and only if $F_{lin}(\lambda_d \circ \rho(N)) = \Theta(F_{lin}(\sigma(N)))$ which holds if and only if $F_{norm}(\lambda_d \circ \rho(N)) = \Theta(F_{norm}(\sigma(N)))$, namely ρ is linear. \square

Theorem 3.3.11 gives us the same optimal decay as in one dimension, which is in contrast to the multidimensional tensor case, where the best we can do is have $d - 1$ extra log factors. We can use this result to cover the two dimensional case in full:

Corollary 3.3.12. *In 2D the optimal decay rate of $(B_f^2(\epsilon), B_{sep}^2)$ is represented by $f(N) = N^{-1}$ (for $\epsilon \in I_{J,\epsilon}$). This optimal decay rate is obtained by using a linear ordering. In fact an ordering ρ of B_f^2 is optimal in lines in 2D if and only if it is linear.*

Proof. Using Lemma 2.2.3 we observe that the decay condition (3.38) holds automatically if $d = 2$. Therefore we may apply Theorem 3.3.11 directly. \square

This result *does not extend to higher dimensions*:

Example 3.3.13. *If we do not have condition (3.38) then our argument can break down very badly: For Haar wavelets we have an explicit formula for the Fourier transform of the one-dimensional mother wavelet,*

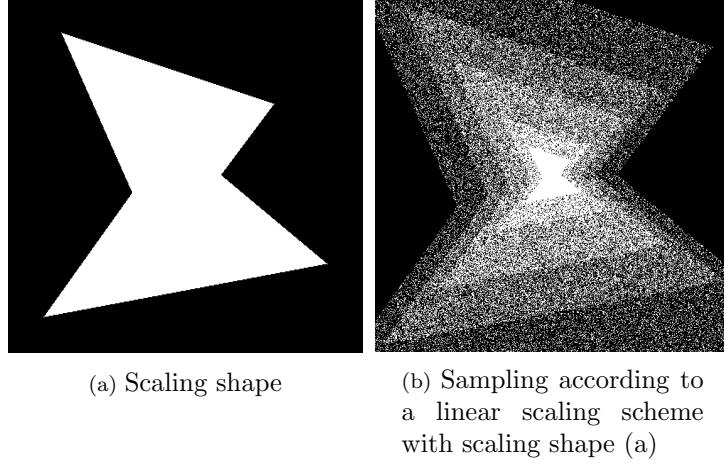
$$\mathcal{F}\phi(\omega) = \frac{\exp(2\pi i\omega) - 1}{2\pi i\omega}.$$

Therefore we have that (3.38) is not satisfied for $d \geq 3$ and furthermore we have (for $\epsilon < 1$ and $J \geq 0$ fixed)

$$|\mathcal{F}\phi(\epsilon 2^{-J}k)| \geq \frac{1}{2\pi\epsilon k}, \quad (3.45)$$

for infinitely many $k \in \mathbb{N}$.

Figure 3.3: A Simple Linear Scaling Scheme



Now consider the case of dD separable Haar wavelets with a linear ordering ρ of the Fourier Basis. Then, for $m \in \mathbb{N}$ such that $\lambda_d \circ \rho(m) = (\lambda_d \circ \rho(m)_1, 0, \dots, 0)$ we know that by (3.45) there are infinitely many m such that

$$\begin{aligned} |\langle \Phi, \rho(m) \rangle|^2 &= \epsilon^d |\mathcal{F}\phi(\epsilon 2^{-J} \lambda_d \circ \rho(m)_1)|^2 \cdot |\mathcal{F}\phi(0)|^{2(d-1)} \\ &\geq \epsilon^d \cdot \frac{1}{(2\pi\epsilon|\lambda_d \circ \rho(m)_1|)^2} \geq \frac{\epsilon^{d-2} E}{4\pi^2 m^{2/d}}, \end{aligned} \quad (3.46)$$

for some constant E using Lemma 3.3.9. Therefore an upper bound of the form $\mathcal{O}(N^{-1})$ is not possible for a linear scaling scheme if $d \geq 3$. This can be rectified by applying a semi-hyperbolic scaling scheme, as in the next subsection.

3.3.4 Examples of Linear Orderings - Linear Scaling Schemes

A wide variety of sampling schemes that are commonly used happen to be linear. In particular we demonstrate that sampling according to how a shape scales linearly from the origin always corresponds to a linear ordering (see Figure 3.3):

Definition 3.3.14. Let $D \subset \mathbb{R}^d$ be bounded with 0 in its interior and define $S_D : \mathbb{Z}^d \rightarrow \mathbb{R}$

$$S_D(x) := \inf \{ \kappa > 0 : x \in \kappa D \}.$$

An ordering $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is said to ‘correspond to a linear scaling scheme with scaling

shape D' if it is consistent with S_D . Furthermore, an ordering $\rho : \mathbb{N} \rightarrow B_f^d(\epsilon)$ is said to ‘correspond to a linear scaling scheme with scaling shape D ’ if it is consistent with $S_D \circ \lambda_d$.

Remark 3.3.15. *If we put a norm $\|\cdot\|$ on \mathbb{Z}^d and take an ordering consistent with this norm then this ordering corresponds to a linear scaling scheme with scaling shape $\{x \in \mathbb{R}^d : \|x\| = 1\}$.*

Lemma 3.3.16. *Let $\rho : \mathbb{N} \rightarrow B_f^d(\epsilon)$ corresponds to a linear scaling scheme with scaling shape D . Then ρ is linear.*

Proof. Let $\sigma = \lambda_d \circ \rho$. Because the scaling shape D is bounded and contains 0 in its interior we have that there exists constants $C_1, C_2 > 0$ such that $C_1 \mathcal{S} \subset D \subset C_2 \mathcal{S}$ where \mathcal{S} is defined to be the unit hypercube, i.e. $\mathcal{S} := \{x \in \mathbb{R}^d : \|x\|_\infty = 1\}$. Therefore if $\|\sigma(N)\|_\infty = L$, then since $D \subset C_2 \mathcal{S}$ we have that $S_D(\sigma(N)) \geq LC_2^{-1}$. Applying this to $C_1 \mathcal{S} \subset D$ we deduce that σ must have enumerated beforehand all points m in \mathbb{Z}^d with $\|m\|_\infty < LC_1 C_2^{-1}$ and there are at least $(2(LC_1 C_2^{-1} - 1) + 1)^d$ of such points. This means that

$$N \geq (2(\|\sigma(N)\|_\infty C_1 C_2^{-1} - 1) + 1)^d \quad \Rightarrow \quad \|\sigma(N)\|_\infty \leq \frac{N^{1/d} + 1}{2C_1 C_2^1} \leq \frac{N^{1/d}}{C_1 C_2^{-1}}, \quad N \in \mathbb{N}.$$

which proves the upper bound. The lower bound is tackled similarly to prove (3.44). \square

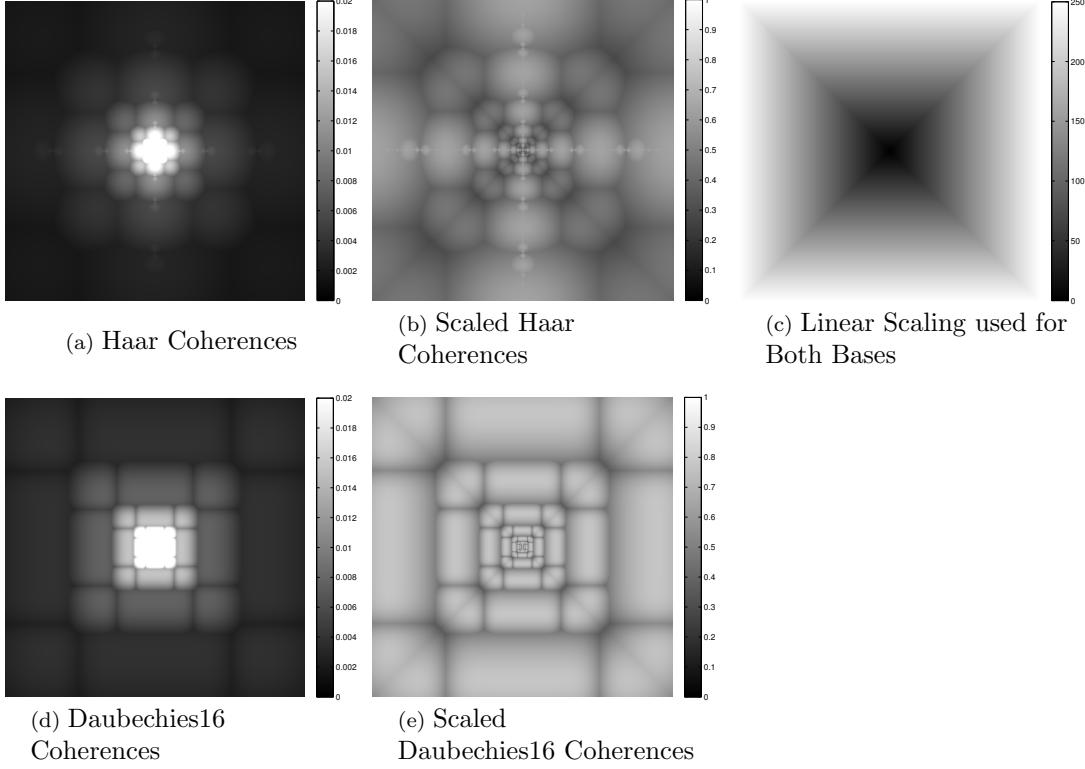
3.3.5 2D Separable Incoherence Plots

By Remark 3.3.8 we have shown that linear orderings are optimal in lines for all 2D Fourier - separable wavelet cases, so this is a good point to have a quick look at a few of these in Figure 3.4.

3.3.6 Semi-Hyperbolic Orderings: Finishing off the Separable Case

By Example 3.3.13 we know that if (3.38) does not hold then our approach of using a linear ordering can fail. We therefore return once more to (3.32). Let us now try to use an approach that is halfway between our two previous linear/hyperbolic approaches. Let $r \in \{1, \dots, d-1\}$ be fixed. We shall first impose a decay condition that is stronger

Figure 3.4: 2D Fourier - Separable Wavelet Coherences



We show the subset $\{-250, -249, \dots, 249, 250\}^2 \subset \mathbb{Z}^2$. Notice again that the scaled coherences are bounded above zero and below 1 indicating that we have characterised the incoherence in terms of the linear scaling used, as shown in Proposition 3.3.7. The incoherences shown in the Figure are square rooted to reduce contrast.

than (2.14) but weaker than (3.38):

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{d/2r}}, \quad \omega \in \mathbb{R} \setminus \{0\}. \quad (3.47)$$

Instead of just taking out the dominant term of the product in (3.32), let us take out the r smallest terms:

$$\begin{aligned} |\langle \Psi_{j,r}^s, \lambda_d^{-1}(n) \rangle|^2 &\leq \epsilon^d 2^{-dj} \cdot \min_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{r=1}^r \frac{K^2}{|\epsilon 2^j n_{i_r}|}^{-d/r} \\ &= K^{2r} \cdot \left(\max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{r=1}^r |n_{i_r}| \right)^{-d/r}, \quad n \in \mathbb{Z}^d, n_i \neq 0, i = 1, \dots, d. \end{aligned} \quad (3.48)$$

Again we can extend this bound to all $n \in \mathbb{Z}^d$:

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 \leq \max(K^{2r}, 1) \cdot \left(\max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{r=1}^r \max(|n_{i_r}|, 1) \right)^{-d/r}. \quad (3.49)$$

We deduce that the function

$$F_{\text{hyp},r} : \mathbb{Z}^d \rightarrow \mathbb{R}, F_{\text{hyp},r}(n) = \left(\max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{r=1}^r \max(|n_{i_r}|, 1) \right)^{-d/r}$$

dominates the optimal decay of $(B_f(\epsilon)^d, B_{\text{sep}}^d)$.

Definition 3.3.17. Let us define, for $r, d \in \mathbb{N}, r \leq d$ the function

$$H_{d,r}(n) := \max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{j=1}^r \max(|n_{i_j}|, 1), \quad n \in \mathbb{Z}^d.$$

Then we say an ordering $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is ‘semi-hyperbolic of order r in d dimensions’ if it is consistent with $H_{d,r}$.

Figure 3.5 presents some isosurface plots of $H_{3,r}$ for the various values of r

Notice that a semi-hyperbolic ordering of order d in d dimensions corresponds to the hyperbolic cross in \mathbb{Z}^d (see Example 3.2.11). Furthermore, if $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is a semi-hyperbolic ordering of order 1 in d dimensions then, by Remark 3.3.15, σ corresponds to a linear scaling scheme because $H_{d,1}(n) = \|n\|_\infty$ for the componentwise max norm $\|\cdot\|_\infty$ on \mathbb{R}^d . Like in the linear and hyperbolic cases discussed in the previous sections, we want to determine how $H_{d,r}(\sigma(n))$ scales with $n \in \mathbb{N}$.

Lemma 3.3.18. 1). Let $r, d \in \mathbb{N}, r \leq d-1$ be fixed. Let us define

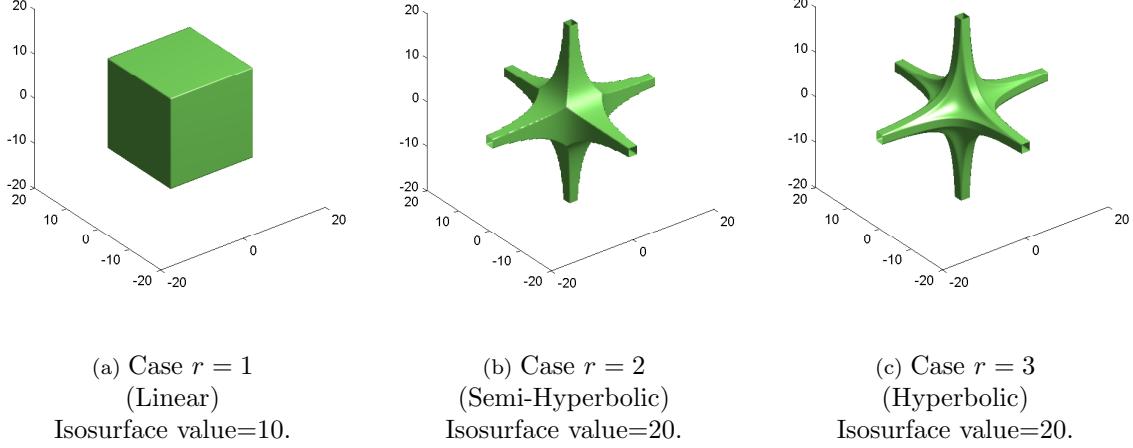
$$S_{d,r}(n) := |\{m \in \mathbb{Z}^d : H_{d,r}(m) \leq n\}|, \quad n \in \mathbb{N}.$$

Then there is a constant $C > 0$ such that

$$n^{d/r} \leq S_{d,r}(n) \leq C \cdot n^{d/r}, \quad n \in \mathbb{N}.$$

2). If $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is semi-hyperbolic of order r with $r \leq d-1$ then

$$H_{d,r}(\sigma(N)) = \Theta(N^{r/d}), \quad N \rightarrow \infty.$$

Figure 3.5: Isosurfaces of $H_{3,r}$, $r = 1, 2, 3$ describing the three types of ordering available in 3D

Proof. 1). For notational simplicity we prove the same bounds but with $S_{d,r}$ replaced by the smaller set

$$\tilde{S}_{d,r}(n) := |\{m \in \mathbb{N}^d : H_{d,r}(m) \leq n\}|, \quad n \in \mathbb{N}.$$

The same bounds for $S_{d,r}$ then follows immediately, albeit with a larger constant $C > 0$. The lower bound is straightforward since the set defining $\tilde{S}_{d,r}(n)$ contains the set $\{m \in \mathbb{N}^d : m_i \leq n^{1/r}, i = 1, \dots, d\}$. We prove the upper bound by induction on r . The case $r = 1$ is clear because $\tilde{S}_{d,1}(n)$ is simply the number of points inside a d -dimensional hypercube with side length n . Suppose the result holds for $r = r' - 1$. We use the following set inclusion:

$$\begin{aligned} \{m \in \mathbb{Z}^d : H_{d,r'}(m) \leq n\} &\subset \{m \in \mathbb{N}^d : m_i \leq n^{1/r'}, i = 1, \dots, d\} \\ &\cup \bigcup_{i=1}^d \{m \in \mathbb{N}^d : n^{1/r'} \leq m_i \leq n, H_{d-1,r'-1}(\tilde{m}_i) \leq n/m_i\}, \end{aligned} \tag{3.50}$$

where \tilde{m}_i here refers to m with the i th entry removed. The cardinality of the first set on the right is just $n^{d/r'}$ and so we are done if we can show that for some constant $C > 0$,

$$|\{m \in \mathbb{N}^d : n^{1/r'} \leq m_1 \leq n, H_{d-1,r'-1}((m_2, \dots, m_d)) \leq n/m_1\}| \leq Cn^{d/r'}, \quad n \in \mathbb{N}$$

We achieve this by applying our inductive hypothesis:

$$\begin{aligned}
& |\{m \in \mathbb{N}^d : n^{1/r'} \leq m_1 \leq n, H_{d-1,r'-1}((m_2, \dots, m_d)) \leq n/m_1\}| \\
& \leq \sum_{i=\lfloor n^{1/r'} \rfloor}^n S_{d-1,r'-1}(\lfloor n/i \rfloor) \leq C' \cdot \sum_{i=\lfloor n^{1/r'} \rfloor}^n (n/i)^{(d-1)/(r'-1)} \\
& \leq C' n^{(d-1)/(r'-1)} \cdot \int_{n^{1/r'}-2}^n x^{-(d-1)/(r'-1)} dx \\
& \leq C' n^{(d-1)/(r'-1)} \cdot (n^{1/r'} - 2)^{(1-(d-1)/(r'-1))}, \quad (\text{noting } r' \leq d-1)
\end{aligned} \tag{3.51}$$

where $C' > 0$ is some constant. We can replace $(n^{1/r'} - 2)$ by $n^{1/r'}$ in the above by changing the constant C' and assuming $n > 2^{r'}$. Finally, we notice that the exponents add to the desired expression:

$$\frac{d-1}{r-1} + \frac{1}{r'} \left(1 - \frac{d-1}{r'-1}\right) = \frac{d-1}{r'-1} - \frac{d-r'}{r'(r'-1)} = \frac{d}{r'}.$$

This gives the required upper bound for $n > 2^{r'}$. Since the terms involved are all positive, we can just increase the constant C' to include the cases $n \leq 2^{r'}$. This shows that the result holds for $r = r'$ and the induction argument is complete.

2.) By consistency we know that

$$S_{d,r}(H_{d,r}(\sigma(n)) - 1) \leq n \leq S_{d,r}(H_{d,r}(\sigma(n))), \quad n \in \mathbb{N}$$

and therefore we can directly apply part 1 to deduce

$$(H_{d,r}(\sigma(n)) - 1)^{d/r} \leq n \leq C \cdot (H_{d,r}(\sigma(n)))^{d/r}, \quad n \in \mathbb{N},$$

from which the result follows. \square

Armed with this result, we can now completely tackle the separable wavelet case.

Theorem 3.3.19. *Suppose that the scaling function ϕ corresponding to the separable wavelet basis B_{sep}^d , satisfies (3.47) for some constant $K \geq 0$ and $r \in \{1, \dots, d-1\}$. Next let $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ be semi-hyperbolic of order r in d dimensions and $\rho := \lambda_d^{-1} \circ \sigma$. Finally, we let $U = [(B_f^d(\epsilon), \rho), (B_{sep}^d, \tau)]$, where τ is an ordering of B_{sep}^d (and $\epsilon \in I_{J,p}$). Then*

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

Furthermore it follows that the ordering ρ is optimal for the basis pair $(B_f^d(\epsilon), B_{sep}^d)$.

Proof. Applying part 1.) from Proposition 3.3.7 (with ϵ fixed) to part 2.) of Lemma 3.1.3 immediately gives us the lower bound for the semihyperbolic ordering since this bound also holds for the optimal decay rate. Furthermore this lower bound holds for any other ordering and therefore if we have the upper bound then the ordering ρ is automatically optimal. We now focus on the upper bound.

By (3.49) we know that the optimal decay of $(B_f^d(\epsilon), B_{sep}^d)$ is dominated by $F_{hyp,r}$. Therefore by part 1.) of Lemma 3.1.3 if $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ is consistent with $1/F_{hyp,r}$, i.e. σ is semihyperbolic of order r then we can bound the row incoherence $\mu(\pi_N U)$ by $F_{hyp,r}(\sigma(N)) = H_{d,r}^{-d/r}(\sigma(N)) = \Theta((N^{r/d})^{-d/r}) = \Theta(N^{-1})$ by Lemma 3.3.18. Since N^{-1} is decreasing this bound extends to $\mu(R_N U)$. \square

Finally we can summarise our results on the $(B_f^d(\epsilon), B_{sep}^d)$ case as follows:

Theorem 3.3.20. *Let ρ be a Linear ordering of the d -dimensional Fourier basis $B_f^d(\epsilon)$ with $\epsilon \in I_{J,p}$, τ a leveled ordering of the d -dimensional separable wavelet basis B_{sep}^d and $U = [(B_f^d(\epsilon), \rho), (B_{sep}^d, \tau)]$. Furthermore, suppose that the decay condition (2.14) holds for the wavelet basis. Then, keeping $\epsilon > 0$ fixed, we have the decay*

$$\mu(\pi_N U), \quad \mu(U\pi_N) = \Theta(N^{-1}) \quad \forall N \rightarrow \infty. \quad (3.52)$$

Let us now instead replace ρ by a semi-hyperbolic ordering of order r in d dimensions with $r \in \{1, \dots, d-1\}$ and assume the weaker decay condition (3.47). Then, keeping $\epsilon > 0$ fixed, we have the decay

$$\mu(R_N U), \quad \mu(U\pi_N) = \Theta(N^{-1}) \quad \forall N \rightarrow \infty, \quad (3.53)$$

and furthermore ρ is optimal for the basis pair $(B_f^d(\epsilon), B_{sep}^d)$. Since, for any separable Daubechies wavelet basis, (3.47) always holds for $r = d-1$ any semi-hyperbolic ordering of order $d-1$ in d dimensions will produce (3.53).

Proof. (3.52) follows from Theorems 3.3.5 and 3.3.11. (3.53) follows from Theorem 3.3.5 and Theorem 3.3.19. To show that (3.53) always holds for a $d-1$ degree semi-hyperbolic ordering in d dimensions, we note that the weakest decay on the scaling function ϕ is $|\mathcal{F}\phi(\omega)| \leq K \cdot |\omega|^{-1}$ (see Lemma 2.2.3) and therefore (3.47) is automatically satisfied for $r = d-1$. \square

3.3.7 Optimal Orderings and Wavelet Smoothness

Theorem 3.3.20 demonstrates how certain degrees of smoothness, in terms of decay of the Fourier transform, allows us to show certain orderings are optimal and this smoothness requirement becomes increasingly more demanding as the dimension increases. But if a certain ordering is optimal for the basis pair $(B_f^d(\epsilon), B_{\text{sep}}^d)$, does this mean that the wavelet must also have some degree of smoothness as well? The answer to this question turns out to be yes, and it is the goal of this section to prove this result.

We shall rely heavily on the following simple result from [45, Thm. 9.4]:

Theorem 3.3.21. *Let⁴ $f : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ be continuous and for $k \in \mathbb{Z}$ define $\hat{f}(k) = \int_0^1 f(x) \exp(2\pi i k x) dx$. If $\sum_{k=-\infty}^{\infty} |k| |\hat{f}(k)| < \infty$ then $f \in C^1$. Consequently, using $\hat{f}'(k) = (2\pi i k)^{-1} \cdot \hat{f}(k)$, if $\sum_{k=-\infty}^{\infty} |k|^n |\hat{f}(k)| < \infty$ then $f \in C^n$.*

Now the main result itself:

Theorem 3.3.22. *Let $\sigma : \mathbb{N} \rightarrow \mathbb{Z}^d$ be semihyperbolic of order $r < d$ in d dimensions and let $\rho := \lambda_d^{-1} \circ \sigma : \mathbb{N} \rightarrow B_f^d(\epsilon)$ where $\epsilon \in I_{J,p}$. Then if ρ is optimal for the basis pair $(B_f^d(\epsilon), B_{\text{sep}}^d)$ then $\phi \in C^l$ for any $l \in \mathbb{N} \cup \{0\}$ with $l + 1 < d/2r$.*

Proof. By Theorem 3.3.20 we know that the optimal decay rate for the basis pair is N^{-1} , therefore if ρ is optimal we must have, for some constant $C_1 > 0$, the bound

$$\sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1} \circ \sigma(N) \rangle|^2 \leq C_1 \cdot N^{-1}, \quad N \in \mathbb{N}.$$

Next since σ is semihyperbolic we also know that, by Lemma 3.3.18, there is a constant $C_2 > 0$ such that

$$H_{d,r}(\sigma(N)) \leq C_2 \cdot N^{r/d}, \quad N \in \mathbb{N}.$$

Consequently we deduce,

$$\begin{aligned} \sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1} \circ \sigma(N) \rangle|^2 &\leq C_1 \cdot N^{-1} \leq C_1 C_2^{-d/r} \cdot H_{d,r}^{-d/r}(\sigma(N)), \quad N \in \mathbb{N}. \\ \Rightarrow \sup_{g \in B_{\text{sep}}^d} |\langle g, \lambda_d^{-1}(n) \rangle|^2 &\leq \frac{C_1 C_2^{-d/r}}{\left(\max_{\substack{i_1, \dots, i_r \in \{1, \dots, d\} \\ i_1 < \dots < i_r}} \prod_{j=1}^r \max(n_{i_j}, 1) \right)^{d/r}}, \quad n \in \mathbb{Z}^d. \end{aligned} \tag{3.54}$$

⁴ \mathbb{R}/\mathbb{Z} denotes the unit circle which we write as $[0, 1)$ with the quotient topology induced by $M : \mathbb{R} \rightarrow [0, 1)$, $M(x) = x(\text{mod } 1)$.

Letting $g = \Psi_{J,0}^s$ where $s = \{0, \dots, 0\}$ and $n = (k, 0, \dots, 0)$ for $k \in \mathbb{Z}$ we see that (3.54) becomes

$$\epsilon^d 2^{-dJ} |\mathcal{F}\phi(2^{-J}\epsilon k)|^2 \leq \frac{C_1 C_2^{-d/r}}{\max(|k|, 1)^{d/r}}, \quad k \in \mathbb{Z}. \quad (3.55)$$

Since the scaling function ϕ has compact support in $[-p+1, p]$ and $\epsilon \in I_{J,p}$, $\phi_{J,0}$ can be viewed as a function on \mathbb{R}/\mathbb{Z} and (3.55) describes a bound on the Fourier coefficients of ϕ . Formally, if we write $\varphi(x) := \phi(2^J \epsilon^{-1}(x - 1/2))$, then since $\epsilon \in I_{J,p}$ we have that φ is supported in $[0, 1]$ and (3.55) becomes, for some constant $D(\epsilon, J, p) > 0$:

$$|\mathcal{F}\varphi(k)|^2 = |\widehat{\varphi}(k)|^2 \leq \frac{D}{\max(|k|, 1)^{d/r}}, \quad k \in \mathbb{Z}.$$

If $\phi \in C^0$ then the result follows from Theorem 3.3.21. If $\phi \notin C^0$, i.e. ϕ corresponds to a Haar wavelet basis, then (3.55) cannot hold with $d/2r > 1$ as this would contradict (3.45). \square

Corollary 3.3.23. *Let the scaling function ϕ corresponding to the Daubechies wavelet basis B_{sep}^d be fixed. Then for every order $r \in \mathbb{N}$, there exists a dimension $d' \in \mathbb{N}, d' > r$ such that for all $d \geq d'$, we have that a semihyperbolic ordering σ of order r in d dimensions is such that $\rho = \lambda_d^{-1} \circ \sigma$ is not optimal for the basis pair $(B_f^d(\epsilon), B_{\text{sep}}^d)$.*

Proof. If this result was not true then we would deduce by Theorem 3.3.22 that the wavelet ϕ satisfies $\phi \in C^\infty$, which is a contradiction because no compactly supported wavelet can be infinitely smooth [22, Thm. 3.8]. \square

3.3.8 Hierarchy of Semihyperbolic Orderings

One other notable point from Theorem 3.3.20 is that we can have multiple values of r such that if σ is semi-hyperbolic of order r in d dimensions then $\rho = \lambda_d^{-1} \circ \sigma$ is optimal for the basis pair $(B_f^d, B_{\text{sep}}^d)$, so which one should we choose? We know that in the case of sufficient smoothness linear orderings are optimal in lines and therefore this suggests that the lower the order r the stronger the optimality result. We now seek to prove this conjecture.

Lemma 3.3.24. *Let $r, r', d \in \mathbb{N}, r \leq r' \leq d$. Then for all $n \in \mathbb{Z}^d$ we have that $H_{d,r'}^r(n) \leq H_{d,r}^{r'}(n)$.*

Proof. Let $n \in \mathbb{Z}^d$ be fixed. For each $j = 1, \dots, d$ let i_j denote the j th largest terms of the

form $\max(|n_{i_j}|, 1)$. Observe that

$$\begin{aligned} H_{d,r}(n) &= \prod_{j=1}^r \max(|n_{i_j}|, 1), \quad H_{d,r'}(n) = \prod_{j=1}^{r'} \max(|n_{i_j}|, 1), \\ &\Rightarrow \frac{H_{d,r}^{r'}(n)}{H_{d,r'}^r(n)} = \frac{\prod_{j=1}^r \max(|n_{i_j}|, 1)^{r'-r}}{\prod_{j=r+1}^{r'} \max(|n_{i_j}|, 1)^r}. \end{aligned}$$

Finally we observe that the numerator and denominator have the same number ($r(r' - r)$) of terms in the product and that each term in the numerator is greater than each term in the denominator, proving the inequality. \square

Corollary 3.3.25. *Let $r, r', d \in \mathbb{N}, r \leq r' < d$ and σ, σ' be semihyperbolic of orders r, r' in d dimensions respectively. If $\rho = \lambda_d^{-1} \circ \sigma$ is optimal for the basis pair $(B_f^d(\epsilon), B_{sep}^d)$ then so is $\rho' = \lambda_d^{-1} \circ \sigma'$.*

Proof. Recalling (3.54) we know that there is a constant $C > 0$ such that

$$\begin{aligned} \sup_{g \in B_{sep}^d} |\langle g, \lambda_d^{-1} \circ \sigma(N) \rangle|^2 &\leq C \cdot H_{d,r}^{-d/r}(\sigma(N)), \quad N \in \mathbb{N}, \\ &\Rightarrow \sup_{g \in B_{sep}^d} |\langle g, \lambda_d^{-1} \circ \sigma'(N) \rangle|^2 \leq C' \cdot H_{d,r'}^{-d/r'}(\sigma'(N)), \quad N \in \mathbb{N}, \end{aligned}$$

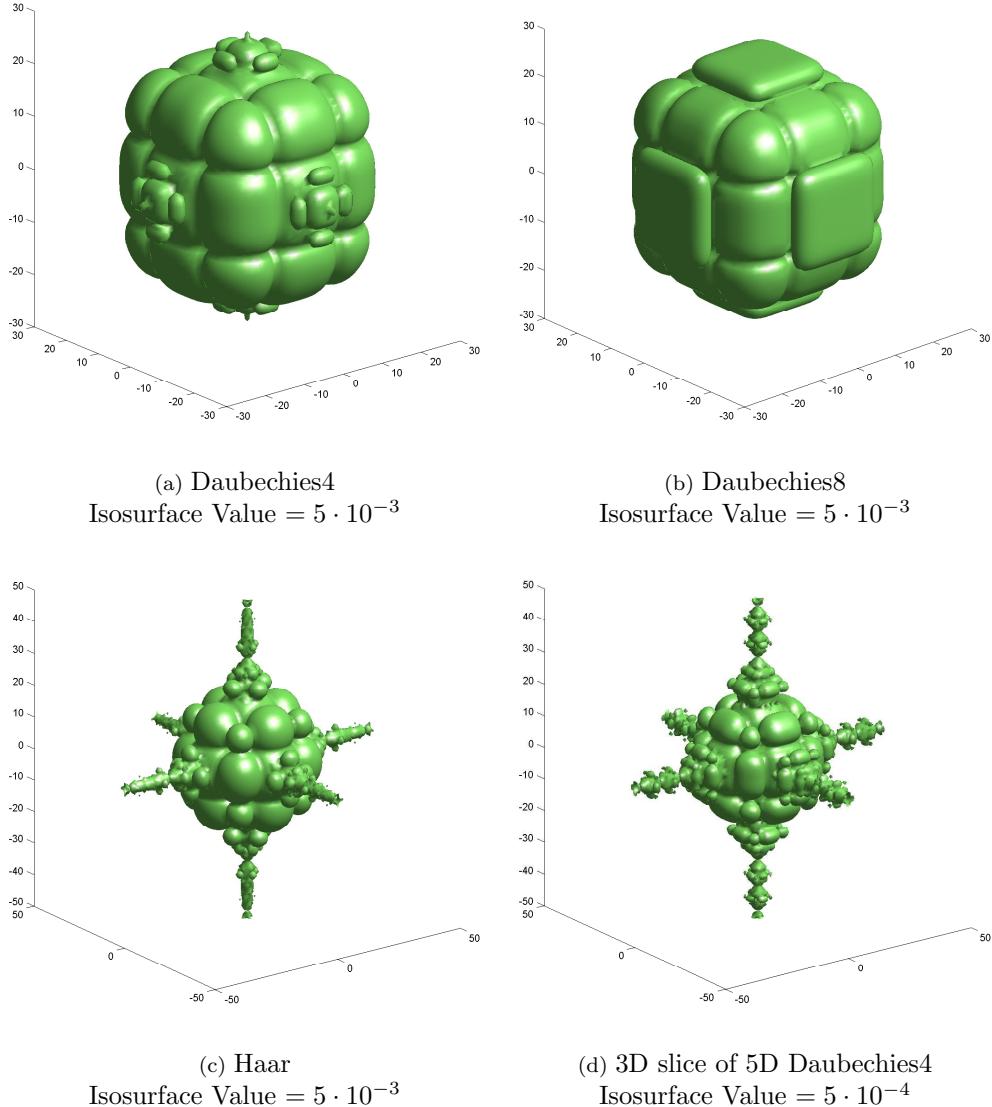
where we have used Lemma 3.3.24 on the second line. We then apply Lemma 3.3.18 to deduce the result. \square

Remark 3.3.26. Corollary 3.3.25 tells us that if there are several orders r that give us optimality then the smallest r possible, say r^* , is the strongest result.

3.3.9 3D Separable Incoherences

We have found optimal orderings for every multidimensional Fourier- separable wavelet case however, we have not shown that (apart from in the linear case with sufficient Fourier decay) that the ordering is optimal in lines and we have not characterized the decay. Therefore it is of interest to see how the incoherence scales in further detail by directly imaging them in 3D. We do this by drawing levels sets in \mathbb{Z}^3 , as seen in Figure 3.6.

Figure 3.6: 3D Fourier - Separable Wavelet Incoherence Isosurface Plots



We draw the isosurface plots over the subset $\{-50, -49, \dots, 49, 50\}^3 \subset \mathbb{Z}^3$. These pictures should be compared with the ordering plots in Figure 3.5. Notice that for the smoother wavelets in (a) & (b), the growth matches that of a linear ordering however the 3D Haar case lacks this smoothness, resulting in semi-hyperbolic scaling in (c). If we keep the wavelet basis fixed and let the dimension increase, the scaling becomes increasingly hyperbolic, as seen in (d) and proved in Corollary 3.3.23.

It is worth mentioning briefly how the 3D coherence data is calculated in these cases. If one tried to calculate these values naively by working with U directly then the finite section of U would have to have at least $100^3 \cdot 100^3 = 10^{12}$ values, which is completely

unfeasible. Fortunately, we only care about the absolute values of this matrix, which means there is considerable redundancy. In fact, this redundancy combined with the tensor product nature of the elements in the separable wavelet basis means that all we really have to calculate are values of the form

$$|\mathcal{F}\psi(\epsilon 2^{-j}m)|, \quad |\mathcal{F}\phi(\epsilon 2^{-j}m)|, \quad k = -50, \dots, 50, \quad j \in \mathbb{N}. \quad (3.56)$$

This is a substantial simplification as around the order of $200 \cdot 7$ Fourier transform values have to calculated versus the previous 10^{12} . The Fourier transfrom values can then be quickly computed using the wavelet filter directly with exponential convergence speeds [22, Prop. 2.21].

3.3.10 Comments on Deriving the $s \log N$ Rate in Multiple Dimensions

Back in Section 2.8, we explained how the classical $s \log(N)$ sampling rate can be derived for the one-dimensional Fourier-wavelet case using the $\Theta(N^{-1})$ decay rates and some extra work involving using the decay of the Fourier transform of the generating wavelet. In particular, in Chapter 2, we needed stronger decay to derive the $s \log(N)$ sampling rate than that of (2.14).

Therefore, extrapolating to the higher dimensional cases, we would expect that we would need stronger decay to derive $s \log(N)$ sampling. However, it turns that only marginally better decay than (3.38) is required to derive the $s \log(N)$ rate for linear orderings, namely

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{d/2+\alpha}}, \quad \omega \in \mathbb{R} \setminus \{0\}, \quad (3.57)$$

for some $\alpha > 0$. The $s \log(N)$ rate can then be proved as follows: split the wavelet basis into levels (2^{dj} in each level j) and split the Fourier basis into the same groupings under a Linear Ordering. Assuming (3.57) holds, we then know that the corresponding change of basis matrix U satisfies $\mu(\pi_N U), \mu(U \pi_N) = \Theta(N^{-1})$. However the tricky part requires proving a version of (2.64). In our case it can be shown that, under (3.57), we have

$$|U_{NM}|^2 \leq C \cdot N^{-1} \min((MN^{-1})^{2\alpha/d}, NM^{-1}), \quad N, M \in \mathbb{N}. \quad (3.58)$$

This is derived by using the bound:

$$\begin{aligned} |U_{NM}|^2 &= |2^{-dj(M)}\epsilon^d \mathcal{F}\psi(\epsilon 2^{-j(M)}\rho \circ \lambda(N))|^2 \\ &\leq 2^{2\alpha j(M)}\epsilon^{-2\alpha} K \|\rho \circ \lambda(N)\|^{-(d+2\alpha)} \\ &= \Theta(M^{2\alpha/d}N^{-(d+2\alpha)/d}) = \Theta(N^{-1}(MN^{-1})^{2\alpha/d}). \end{aligned} \quad (3.59)$$

From here the following bound on the local coherence can be derived:

$$\mu(P_{N_{k-1}}^{N_k} UP_{M_{l-1}}^{M_l}) \lesssim 2^{-dk} 2^{-2\alpha|k-l|},$$

and the same geometric trick can be used to get the $s \log(N)$ rate using (1.25).

One point to notice is that (3.57) does exclude the 2D separable Haar case, which leads us to conclude that its use should be avoided in 2+ rather than 3+ dimensions as was suggested by Example 3.3.13.

3.4 Asymptotic Incoherence and Two Dimensional Compressed Sensing

Consider the problem of reconstructing a function $f \in L^2([0, 1]^2)$ from its samples $\{\langle f, g \rangle : g \in B_f^2(1)\}$. The function f is reconstructed as follows: Let

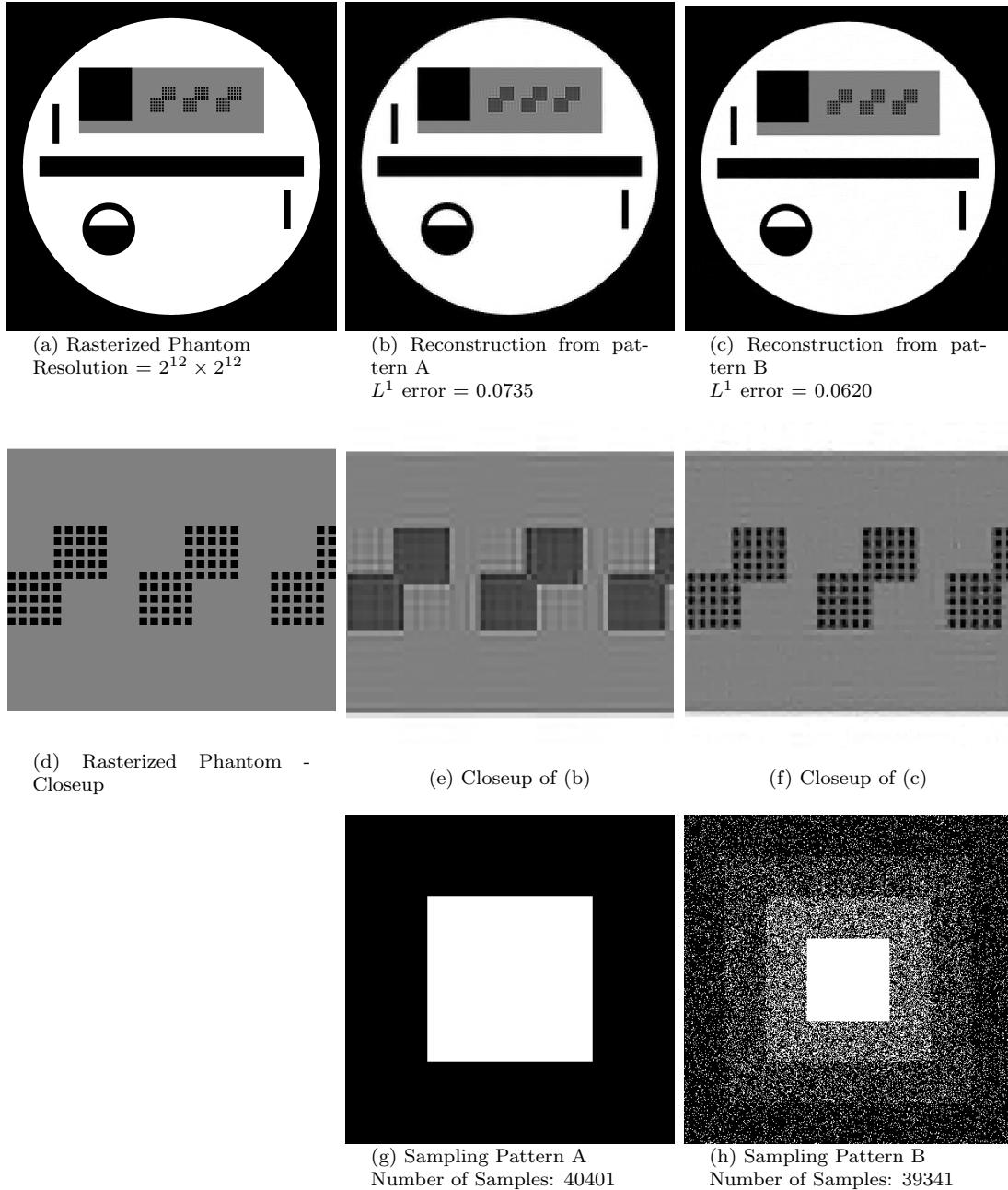
$$U := [(B_f^2(1), \rho), (B_2, \tau)]$$

for some orderings ρ, τ and a reconstruction basis B_2 . Next let $\Omega \subset \mathbb{N}$ denote the set of subsamples from $B_f^2(1)$ (indexed by ρ), P_Ω the projection operator onto Ω and $\hat{f} := (\langle f, \rho(m) \rangle)_{m \in \mathbb{N}}$. We then attempt to approximate f by $\sum_{n=1}^{\infty} \tilde{x}_n \tau(n)$ where $\tilde{x} \in \ell^1(\mathbb{N})$ solves the optimisation problem

$$\min_{x \in \ell^1(\mathbb{N})} \|x\|_1 \quad \text{subject to} \quad P_\Omega U x = P_\Omega \hat{f}. \quad (3.60)$$

Since the optimisation problem is infinite dimensional we cannot solve it numerically so instead we proceed as in [27] and truncate the problem, approximating f by

Figure 3.7: Simple Resolution Phantom Experiment



Samples are from the subset $\{-200, -199, \dots, 199, 200\}^2 \subset \mathbb{Z}^2$. Notice that the checkerboard feature are captured by the leveled sampling pattern but not by pattern (a), even though it uses fewer samples. Reconstructions are at a resolution of $2^{10} \times 2^{10}$.

$\sum_{n=1}^R \tilde{x}_n \tau(n)$ (for $R \in \mathbb{N}$ large) where $\tilde{x} = (\tilde{x}_n)_{n=1}^R$ now solves the optimisation problem

$$\min_{x \in \mathbb{C}^R} \|x\|_1 \quad \text{subject to} \quad P_\Omega U P_R x = P_\Omega \hat{f}. \quad (3.61)$$

Like in the previous numerical examples we shall be using the SPGL1 package [12, 13] to solve (3.61).

3.4.1 Demonstrating the Benefits of Multilevel Subsampling

We shall first demonstrate directly how subsampling in levels is beneficial in situations with asymptotic incoherence ($\mu(R_N U) \rightarrow 0$) but poor global incoherence ($\mu(U)$ is relatively large). The image f that we will attempt to reconstruct is made up of regions defined by Bezier curves with one degree of smoothness, as in [36]. This image is intended to model a resolution phantom⁵ which is often used to calibrate MRI devices [46]. A rasterization of this phantom is provided in image (a) of Figure 3.7.

We reconstruct with 2D separable Haar wavelets, ordered according to its resolution levels, from a base level of 0 up to a highest resolution level of 8. The Fourier basis is ordered by the linear consistency function $H_{2,1}$ from Definition 1.7.16, which gives us a square leveling structure when viewed in \mathbb{Z}^2 . We choose these orderings because we know that they are both optimal in lines for the corresponding bases, and therefore should allow reasonable degrees of subsampling when given an (asymptotically) sparse problem.

By looking at Figure 3.7, we observe that subsampling in levels (pattern (b)) allows to pick up features that would be otherwise impossible from a direct linear reconstruction from the first number of samples (pattern (a)) and moreover the L^1 error is smaller.

3.4.2 Tensor vs Separable - Finding a Fair Comparison

We would like to study how different asymptotic incoherence behaviours can impact how well one can subsample. In 2D it would be unwise to compare 2 different separable wavelet bases, since we know that they have the same optimal orderings and decay rates in 2D (see Theorem 3.3.11). Therefore we are left with comparing a separable wavelet basis to a tensor basis. The incoherence decay rates for the 2D Haar cases are shown in the table below for Linear and Hyperbolic orderings of the Fourier basis B_f^2 :

Observe that for linear orderings, there is a large discrepancy between the decay rates, however they are the same for hyperbolic orderings. Therefore, comparing separable and tensor reconstructions appears to be a good method for testing the behaviour of differing speeds of asymptotic incoherence.

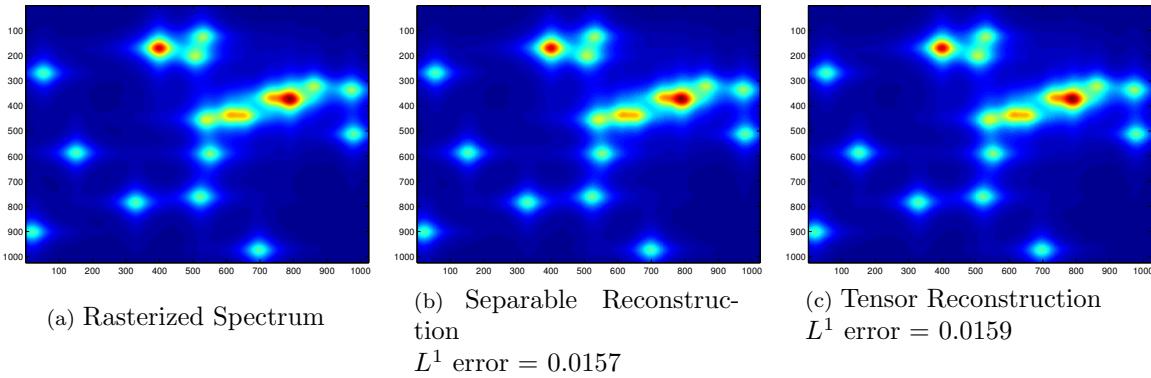
⁵'resolution' here refers to 'resolving' a signal from a MRI device.

2D Haar Basis Incoherence Decay Rates		
Ordering	Tensor	Separable
Linear	$N^{-1/2}$	N^{-1}
Hyperbolic	$\log(N + 1) \cdot N^{-1}$	$\log(N + 1) \cdot N^{-1}$

Table 3.1: The decay rates for the hyperbolic case comes from Theorem 3.2.9 and Proposition 3.3.6. For the linear case, the separable result comes from Theorem 3.3.20 and the tensor result can be deduced from Lemma 3.3.9 applied to (3.22), although we do not provide the details here.

However, there is one serious problem, namely the choice of image f that we would like to reconstruct. Recall from (1.25) that the ability to subsample depends on both the coherence structure of the pair of bases and the sparsity structure of the function f we are trying to reconstruct. Ideally, to isolate the effect of asymptotic incoherence we would like to choose an f that has the same sparsity structure in both a tensor and separable wavelet basis. If f was chosen to be the resolution phantom like before then the tensor wavelet approximation would be a poor comparison to that of the separable wavelet reconstruction (due to a poor resolution structure). Therefore we need to choose a function that we expect to reconstruct well in tensor wavelets, for example a tensor product of one dimensional functions.

Figure 3.8: Spectrum Model and ‘Full Sampling’ Reconstructions



Reconstructions uses all samples from the subset $\{-200, -199, \dots, 199, 200\}^2 \subset \mathbb{Z}^2$. Images are at a resolution of $2^{10} \times 2^{10}$. Haar wavelets are used for tensor and separable cases. Observe that both reconstructions match the original very closely and have similar L^1 approximation errors.

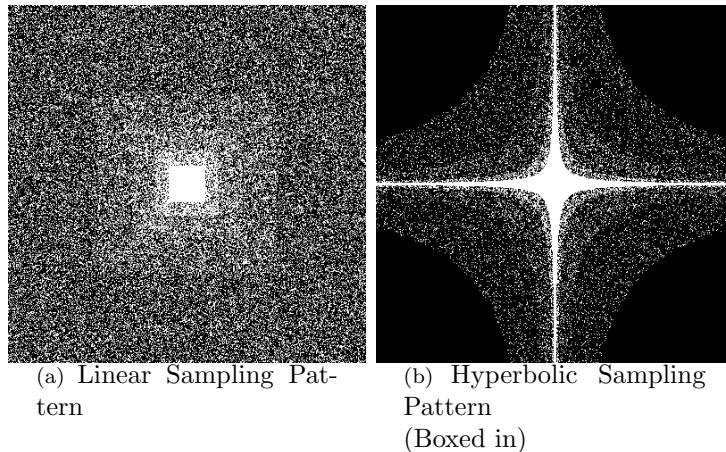
Such an example is provided by NMR spectroscopy [16, Eqn. (5.24)]. A 2D spectrum

is sometimes modelled as a product of 1D Lorentzian functions:

$$\begin{aligned} f(x) &= \sum_{i=1}^r L_{2,p(i),s(i)}(x), \quad x, p(i), s(i) \in \mathbb{R}^2, \\ L_{2,p,s}(x) &= L_{p_1,s_1}(x_1) \cdot L_{p_2,s_2}(x_2), \quad x, p, s \in \mathbb{R}^2 \\ L_{p,s} &= \frac{s}{s^2 + (x - p)^2}, \quad x, p, s \in \mathbb{R}. \end{aligned} \tag{3.62}$$

We consider a specific spectrum f of the above form. By looking at Figure 3.8 we observe that, without any subsampling from the subset $\{-200, -199, \dots, 199, 200\}^2 \subset \mathbb{Z}^2$, the tensor and separable Haar wavelet reconstructions have almost identical L^1 errors, suggesting that this problem does not bias either reconstruction basis. We order the tensor and separable reconstruction bases using their corresponding level based orderings, which are defined in Lemma 3.2.14 and Definition 3.3.3 respectively. For separable wavelets we start at a base level of $J = 0$ and stop at level 8 (so we truncate at the first $2^{10} \times 2^{10}$ wavelet coefficients) and for tensor wavelets we start at level $J = 0$ and stop at level 10 (when the problem was truncated at higher wavelet resolutions the improvement in reconstruction quality was negligible).

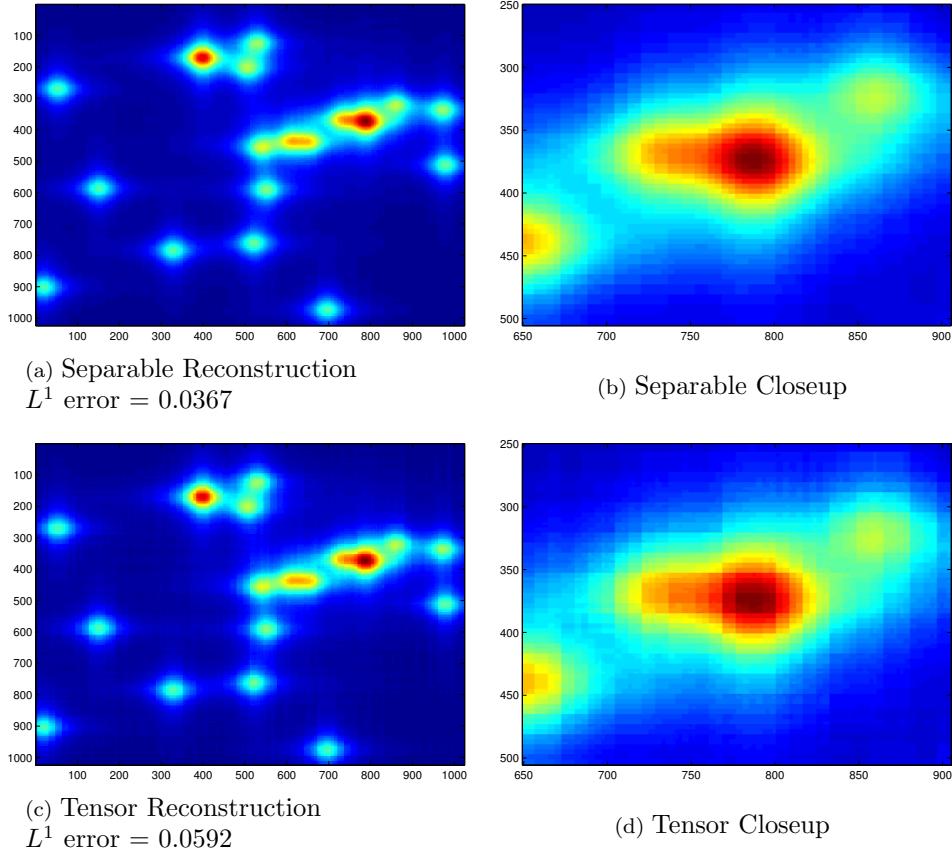
Figure 3.9: Sampling Patterns



Samples are from the subset $\{-200, -199, \dots, 199, 200\}^2 \subset \mathbb{Z}^2$. White indicates sample is taken.

We are now going to test how well these two bases perform under subsampling with different orderings of B_f^2 . Two subsampling patterns, one based on a linear ordering and another on a hyperbolic ordering, are presented in Figure 3.9. Ideally the hyperbolic subsampling pattern would not be restricted by the $\{-200, -199, \dots, 199, 200\}^2$ but this is numerically unfeasible.

Figure 3.10: Reconstructions from Linear Sampling Pattern



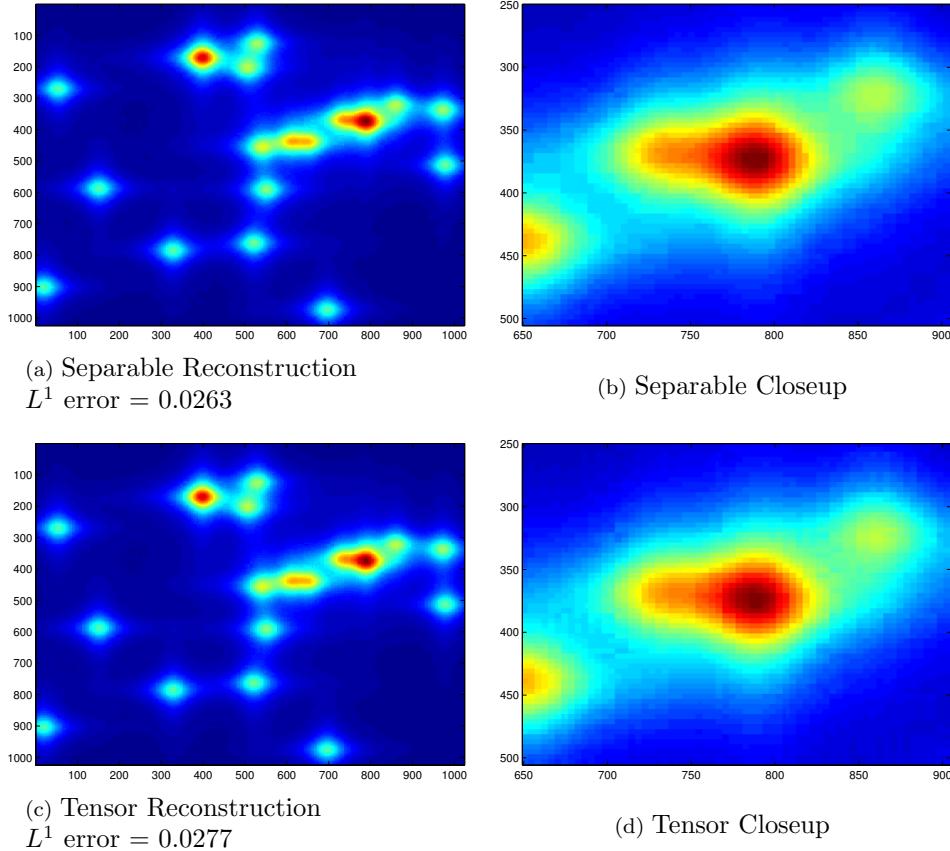
Let us first consider what happens when using pattern (a) (see Figure 3.10). Notice that the separable reconstruction performs far better than the tensor reconstruction and therefore is more tolerant to subsampling with a linear ordering than the tensor case. This is unsurprising as the tensor problem suffers from noticeably large $1/\sqrt{N}$ incoherence when using a linear ordering when compared to the $1/N$ separable decay rate.

Of course we should have fully considered the sparsity of these two problems which also factor into the ability to subsample, however f was specifically chosen because it was sparse in the tensor basis and moreover we have seen that it provides a comparable reconstruction to the separable case when taking a full set of $\{-200, -199, \dots, 199, 200\}^2$ samples.

Next we observe what happens when using the pattern (b) (Figure 3.11). There is now a stark contrast to the linear case, in that both separable and tensor cases provide very similar reconstructions and furthermore the L^1 errors are very close. This suggests

that both problems have similar susceptibility to subsampling when using hyperbolic sampling, which is reflected by their identical rates of incoherence decay with hyperbolic orderings.

Figure 3.11: Reconstructions from Hyperbolic Sampling Pattern



Chapter 4

Compressed Sensing of Helium Atom Scattering Spectra

This Chapter covers an ongoing collaboration between the Surfaces, Microstructure and Fracture (SMF) research group at the Cavendish site and the Applied Functional and Harmonic Analysis (AFHA) research group at DAMPT. This collaborative work implemented both traditional and continuous CS methods for the purposes of reconstructing Helium Atom Scattering (HAS) spectra with the new power supplies at the Cavendish site.

The HAS technique Helium Spin-Echo (HeSE) is an ideal tool for probing the surface layer of crystal structures. Not only can HeSE be used to study static crystal structures in the manner of Bragg's original work [47], but it can also be used to study dynamic behaviour similar to techniques in the field of neutron scattering [48]. This could be adatoms diffusing on the surface of a rigid structure [49], phonon vibrations [50], etc. Thanks to the work of van Hove [51] et al. on the theory of neutron scattering, a theoretical framework for HeSE was developed as a surface analogue of this earlier work. This framework relates time and position (\mathbf{R}, t) to momentum and energy transfer ($\Delta\mathbf{K}, \Delta E$) as a Fourier pair.

Compressed sensing has long been associated with Nuclear Magnetic Resonance (NMR) based applications such as Magnetic Resonance Imaging [52] and NMR Spectroscopy [17]. While there are clear similarities with these fields, such as Fourier transforms arriving naturally in data acquisition, there are considerable differences in how data is acquired and then processed. For example, in NMR-based experiments the smallest group of data that is taken in one measurement is typically a line (or path) of data

points [53], however in HeSE the smallest group is a single point.

Naturally, HeSE has its own unique set of difficulties that have to be overcome. On the one hand there is the issue of handling projections using the Fourier slice theorem. On the other hand using certain discretisation methods for handling the Fourier transform can lead to non-uniform spacing of data points owing to non-linear transforms that are applied after reconstruction. Recent work by Adcock, Hansen, Poon and Roman [27, 15] overcomes this problem by working with function approximations, avoiding the use of a grid entirely. Furthermore, this theory adds justification to the non-uniform sampling patterns that empirically outperform uniform-sampling techniques that are the norm within conventional compressed sensing.

The goals of this chapter are twofold:

- To outline some of the mathematical models that underpin experimental Helium Spin-Echo and explain the principles behind current compressed sensing techniques.
- To describe how these techniques are applied to HeSE and demonstrate their benefits with real examples of polarisation data sets. This includes a unified approach for simultaneous detection of phonon and diffusive data.

A summary of the type of problems typical to HeSE can be found in Figure 4.1. This demonstrates the various steps that are involved in the entire process from the data sampling in polarisation to data analysis. This underlying physics of this problem is inherently two-dimensional however there are tools for simplifying the problem (such as the Fourier slice theorem 4.2.4) that allow this problem to be treated as a one-dimensional one. This is the approach that is currently adopted by the SMF group however there are plans to treat to full two-dimensional problem where CS stands to make the greatest benefit (see Section 4.4).

4.1 Structure of this Chapter

This Chapter consists of three major parts. Section 4.2 covers the background of the polarisation measurement stage and some background regarding phonons (corresponding to the first arrow in Figure 4.1). After this background, Section 4.3 cover the application of discrete and continuous CS to this stage. Section 4.4 covers the other two stages in Figure 4.1 and covers how continuous CS can be used to benefit the data processing

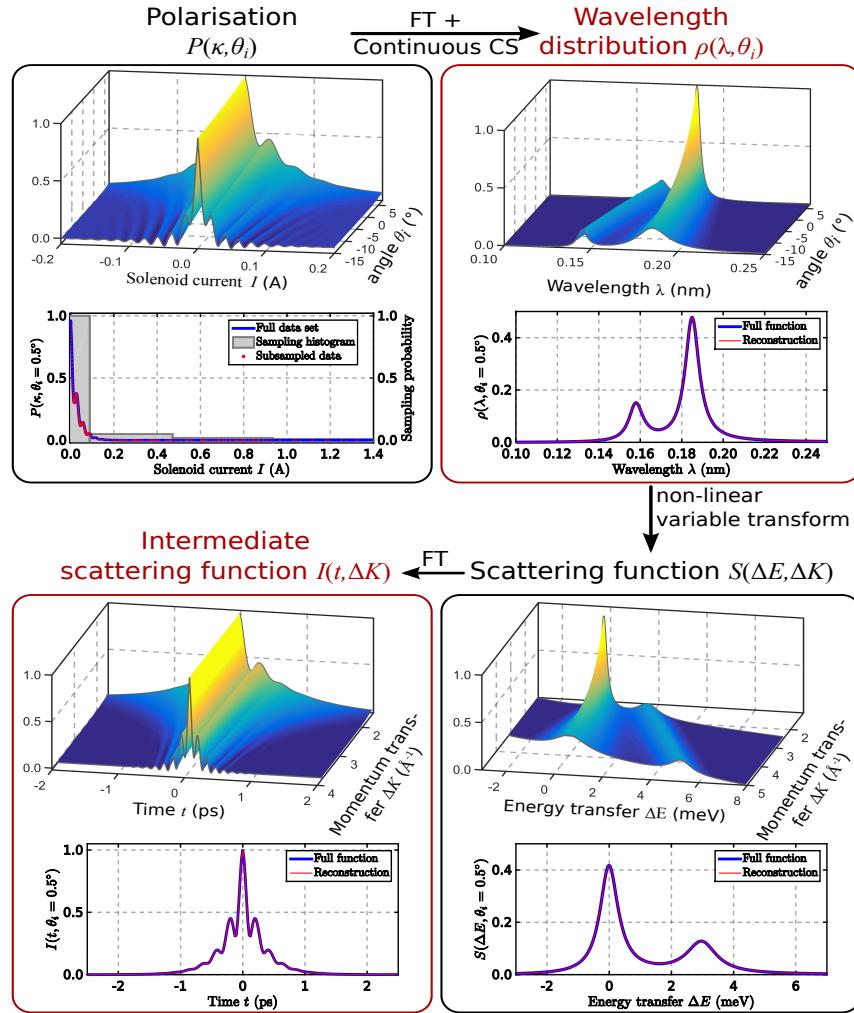


Figure 4.1: Diagram outlining the various stages of data transformation from measurement in polarisation to the ISF describing molecular processes. The upper plots denote the full 2D data while the lower plots are 1D projections/slices. The stages highlighted in red correspond to the target data we wish to reconstruct. The sampled intensities are shown in normalised arbitrary units (a.u.). Figure taken from collaborative work [3] with the SMF group.

between these stages and suggests various alternatives for reconstructing the scattering functions. Chapter 5 also outlines possible routes for the future of the collaboration and assesses the current state of development.

4.2 Background of Helium Spin-Echo Part 1 - Measurements and Phonons

Surface-related phenomena are prevalent throughout everyday life through processes such as friction, corrosion and tension. When designing new materials, such as for

catalysis [54] or electronics [55, 56], one would like to study various vibrational and electronic properties to determine their potential benefits and drawbacks. Neutrons and X-rays have been successful in measuring such phenomena over the entire bulk of a sample. However, to truly understand surface processes one needs to investigate the differences between the structure of a surface and the bulk [57]. Therefore, a purely surface sensitive approach is highly desirable. The repulsive part of the ${}^3\text{He}/\text{surface}$ interaction potential prevents the He atoms from penetrating into the surface layer of materials. Consequently they are ideal as a surface scattering probe.

The basic setup of the Helium spin-echo apparatus at Cambridge's Cavendish laboratory is given in Figure 2. A beam of thermal ${}^3\text{He}$ is generated from the source in a fixed direction. The nuclear spins are polarised and then rotated by a solenoid before being scattering upon the target crystal surface. Afterwards any scattered He atoms heading in the direction of the detector are then rotated by a second solenoid and passed through another polarisation filter. A schematic for the apparatus can be found in [58].

Helium Spin-Echo Spectroscopy should not be confused with Helium Atom Time-of-Flight Spectroscopy [50, 59] where the velocities of the incoming / outgoing Helium atoms are measured directly. In this case there is no Fourier transform in the sampling stage which prevents the application of the CS techniques presented in this chapter. In addition, while the time of flight technique has provided many useful measurements in past decades [60], the Helium Spin-Echo technique benefits from greater energy resolution [19].

Key variables that the operator can freely adjust include

- The currents I_i, I_f that run through Solenoids A and B respectively
- The scattering geometry, in particular the angles θ_i, θ_f made between the surface normal and the surface-to-source/detector directions (see Figure 4.4)

The incoming monochromatic He beam can be modelled as a plane wave [61] with propagation wavevector $\mathbf{k} \in \mathbb{R}^3$ and angular frequency ω :

$$\psi(\mathbf{r}, t) = \exp(i(\mathbf{k} \cdot \mathbf{r} - \omega t)), \quad \mathbf{r} \in \mathbb{R}^3, \quad t \in \mathbb{R}. \quad (4.1)$$

Here, \mathbf{r} denotes position and t is time. The wavevector \mathbf{k} and wavelength λ are related to particle's momentum \mathbf{p} by the de Broglie relations

$$\mathbf{p} = \hbar\mathbf{k}, \quad |\mathbf{p}| = p = \frac{2\pi}{\lambda}. \quad (4.2)$$

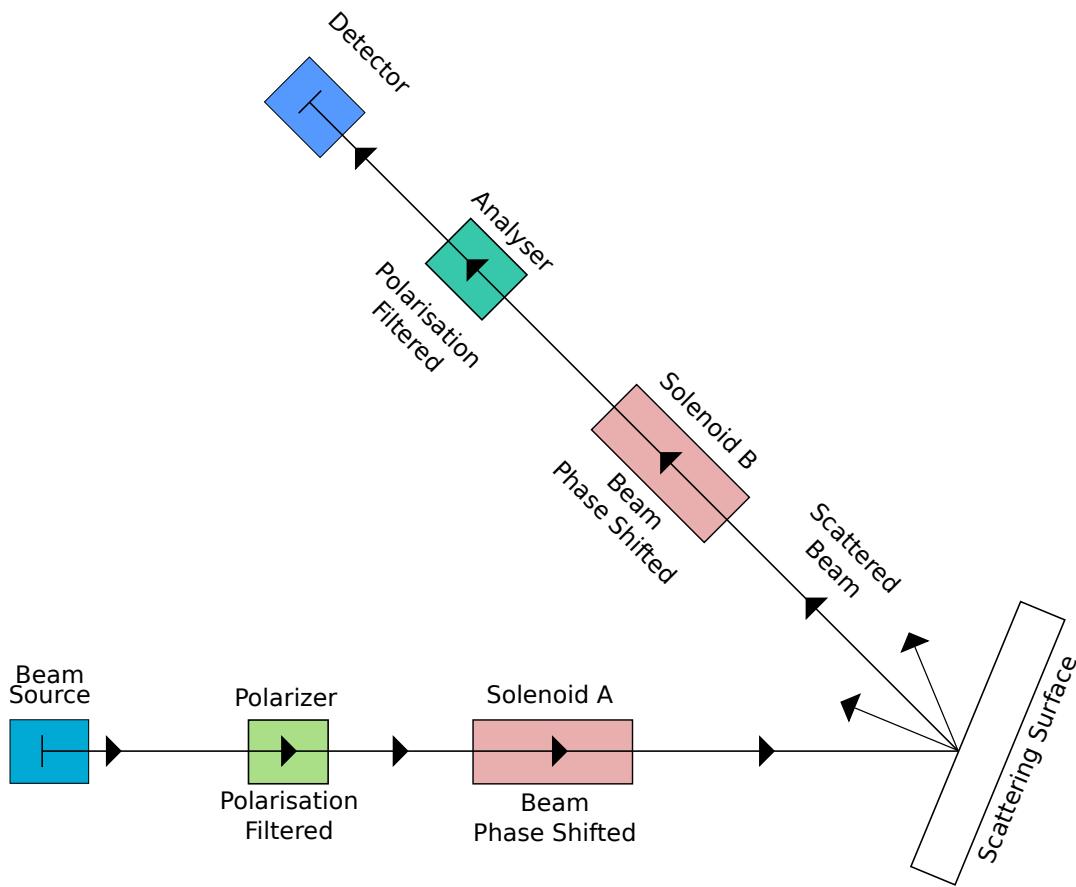


Figure 4.2: Outline of the Cambridge Spin-Echo Apparatus.

Furthermore, the frequency ω is related to the particle energy E by the relation

$$E = \hbar\omega. \quad (4.3)$$

Using these relations we can treat \mathbf{k} as representing momentum and ω as energy. Notice that by Formula (4.1) we have identified two Fourier pairs $(\mathbf{k}, \mathbf{r}), (\omega, t)$. However, at present these pairs only relate to the beam of Helium and not the crystal surface that it will be scattering upon.

4.2.1 Bravais Crystals, Laue Equations and the Reciprocal Lattice

For now suppose that the nuclei in the crystal is static and can be expressed as a Bravais lattice as shown in Figure 4.3. A Bravais crystal is defined to be one whose

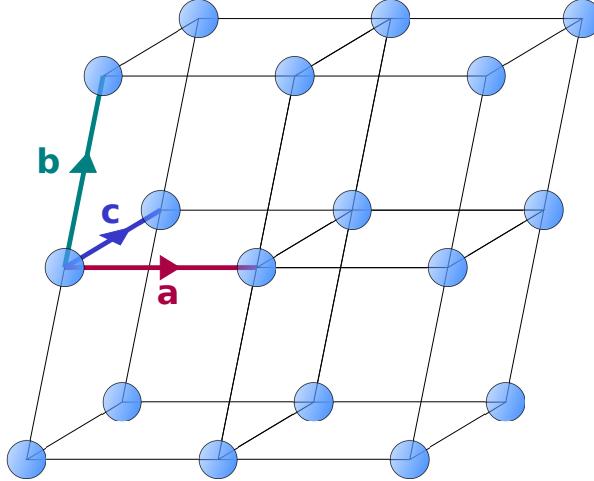


Figure 4.3: A Bravais crystal with the three primitive vectors labelled

nuclei positions are of the form

$$L_{\mathbf{a}, \mathbf{b}, \mathbf{c}} := \{\mathbf{r}_{n_1, n_2, n_3} = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c} : n_1, n_2, n_3 \in \mathbb{Z}\}, \quad \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3. \quad (4.4)$$

We assume that the primitive vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent. Suppose we have a beam of atoms (for example ${}^3\text{He}$) that is incident upon the crystal surface with incident wavevector \mathbf{k}_{in} . The Laue Equations [62] state that the only observable wavevectors \mathbf{k}_{out} that appear in the scattered beam must satisfy (for $\Delta\mathbf{k} = \mathbf{k}_{\text{out}} - \mathbf{k}_{\text{in}}$)

$$\exp(i\Delta\mathbf{k} \cdot \mathbf{r}) = 1, \quad \mathbf{r} \in L_{n_1, n_2, n_3}. \quad (4.5)$$

The set of all vectors $\Delta\mathbf{k}$ that satisfy (4.5) is called the *reciprocal lattice* of the Bravais lattice defined by (4.4), which can be shown to be another 3D Bravais lattice with primitive vectors

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}. \quad (4.6)$$

It is easily verified that the reciprocal lattice of the reciprocal lattice generated by (4.6) is again the original Bravais lattice (4.4). Therefore, if one can find all possible scattered wavevectors \mathbf{k}_{out} then the reciprocal lattice is known and therefore so is the lattice structure of the Bravais crystal.

While this is useful for understanding the basic relationship between wavevectors / momentum and the crystal structure, in practice there are several issues with this

approach, including:

1. We have assumed the nuclei in the crystal are static. Studying dynamic phenomena requires a different approach.
2. The He³ beam probes only the surface layer [63, Pg. 78] so instead of periodicity in three dimensions we are dealing with periodicity in two dimensions.

Over the next few sections, we shall be looking at how these points are dealt with.

4.2.2 Phonons and Changes in Helium Energy

Nuclei positions in a crystal structure undergo vibrations at various modes. The momenta associated with the vibrational modes can be broken down in terms of the three reciprocal lattice vectors:

$$\mathbf{q} = \zeta_1 \mathbf{a}^* + \zeta_2 \mathbf{b}^* + \zeta_3 \mathbf{c}^*, \quad \zeta_i \in (-1/2, 1/2], \quad i = 1, 2, 3. \quad (4.7)$$

In reality the set of modes (4.7) is discrete but over a very fine scale. The above set of vectors is often referred to as the *Brillouin Zone*. Let $u_{\mathbf{n}}$ denote the displacement from equilibrium of the nuclei at equilibrium position $\mathbf{r}_{\mathbf{n}}$ described by (4.4). Then a model for $u_{\mathbf{n}}(t)$ is given by [48, Eq. G.14]:

$$u_{\mathbf{n}}(t) = \sum_{\mathbf{q}, j} \exp(i\mathbf{q} \cdot \mathbf{r}_{\mathbf{n}}) (A_{\mathbf{q}, j} \exp(i\omega_{\mathbf{q}, j} t) + B_{\mathbf{q}, j} \exp(-i\omega_{\mathbf{q}, j} t)) \cdot \mathbf{e}_{\mathbf{q}, j}, \quad (4.8)$$

where the sum runs over all possible wavevectors \mathbf{q} in the Brillouin zone (4.7) and dimensions $j = 1, \dots, 3$, $\mathbf{e}_{\mathbf{q}, j}$ denotes the direction of displacement (called the *polarisation vector*), $\omega_{\mathbf{q}, j}$ refers to the frequency of the vibrations and $A_{\mathbf{q}, j}, B_{\mathbf{q}, j}$ are constants. Although this involves complex quantities, $u_{\mathbf{n}}(t)$ is always real as it is a physical quantity. The frequency $\omega_{\mathbf{q}, j}$ depends on the wavevector \mathbf{q} by a *dispersion relation*. The main property of this relation is that it is periodic over the Brillouin zone, namely that if \mathbf{q}' denotes any reciprocal lattice vector then

$$\omega_{\mathbf{q}+\mathbf{q}', j} = \omega_{\mathbf{q}, j}, \quad \forall \mathbf{q}. \quad (4.9)$$

Like how the particles in the Helium beam can be treated as a wave, the vibrational waves through the crystal can be quantised as particles with their own energies and

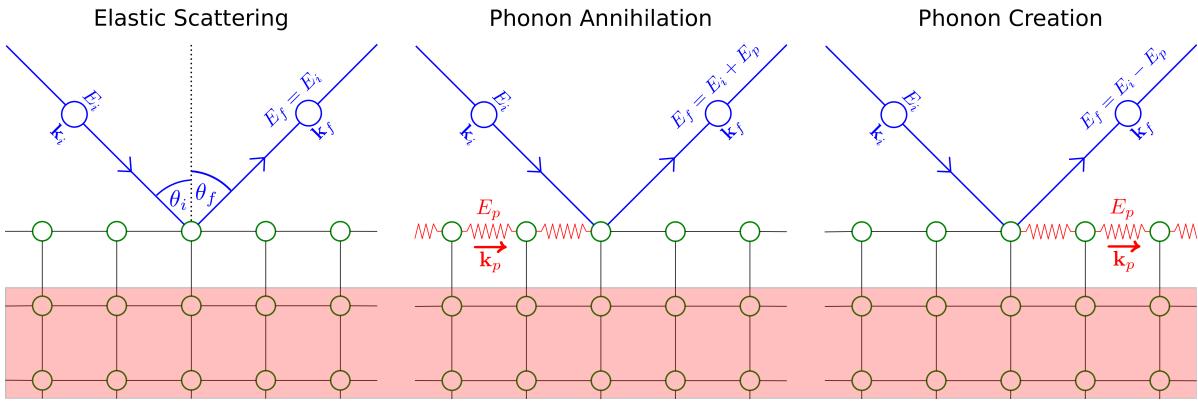


Figure 4.4: Three possible scattering events. The He^3 particle (in blue) can gain or lose energy after scattering. The He^3 particle can only interact with the surface layer and therefore it can only interact with phonons on the surface. The red zone denotes the forbidden region for the He^3 beam.

momenta, which are again described by the Broglie relation (4.2) and equation (4.3). Therefore the wavevector \mathbf{q} can be interpreted as momentum and the frequency $\omega_{\mathbf{q},t}$ as energy. Consider a term in the sum (4.8):

$$\exp(i\mathbf{q} \cdot \mathbf{r}_n)(A_{\mathbf{q},j} \exp(i\omega_{\mathbf{q},j}t) + B_{\mathbf{q},j} \exp(-i\omega_{\mathbf{q},j}t)) \cdot \mathbf{e}_{\mathbf{q},j}, \quad (4.10)$$

Although this quantity has a momentum/wavevector \mathbf{q} , it is only unique up to a reciprocal lattice vector just like (4.9). Therefore, the vibrational mode corresponding to the term (4.10) could have several possible momenta represented by one unique vector \mathbf{q} in the Brillouin zone.

Now we look at how these phonons can influence the scattering of the Helium beam. Figure 4.4 shows three possible outcomes of a single Helium particle scattering upon a crystal surface (more complex interactions are possible [58]). The figure shows a cross section of the crystal where the path of the Helium beam lies in the same cross section, therefore, we are effectively working with a two dimensional problem. We shall assume there is conservation of energy before and after scattering and that the momentum is conserved parallel to the surface of the crystal¹. If we let little i, f denote the incident and final Helium states, little p a possible phonon, \mathbf{k} a wavevector and $k = |\mathbf{k}|$ its magnitude, then by conservation of energy and momentum

$$E_f = E_i + \Delta E, \quad K_i = K_f + \Delta K, \quad (4.11)$$

¹Recall that the Helium beam cannot penetrate below the first layer of the crystal and so conservation of perpendicular momentum is of no use here.

where $K_i = k_i \sin \theta_i, K_f = k_f \sin \theta_f$ denotes the projections of the incident and final momentums $\mathbf{k}_i, \mathbf{k}_f$ onto the surface. The change in energy ΔE is equal to the change in kinetic energy of the Helium particle:

$$\Delta E = \frac{\hbar^2}{2m} k_f^2 - \frac{\hbar^2}{2m} k_i^2. \quad (4.12)$$

In the case of elastic scattering ($\Delta E = 0$) we have $k_i = k_f$ and $\Delta K = G$ where G corresponds to any reciprocal lattice vector parallel to the surface and the cross section, which is consistent with the Laue Equations described in the previous section.

Suppose now that a Helium particle collides with a surface phonon with energy $\Delta E = \hbar\omega_p$. The momentum of the phonon has the form $k_p + G_p$ where \mathbf{k}_p lies in the Brillouin zone and \mathbf{G}_p is a reciprocal lattice vector parallel to the surface. Therefore we expect $\Delta K = k_p + G$ with G as in the elastic case. Similar results hold for phonon creation.

The key point of these calculations is that the changes of energy and momentum contain information about the phonons on the crystal surface.

4.2.3 Solenoid Currents and Spin Phase

We now turn to how the scattering apparatus shown in Figure 2 can be used to measure properties of the He beam. In particular, we show how the solenoid currents (I_i, I_f) and scattering wavelengths (λ_i, λ_f) share a direct Fourier relationship. This section follows closely the review [19] of Alexandrowicz and Jardine.

Recall that we have two solenoids that generate magnetic fields which rotates the polarisation of the ${}^3\text{He}$ beam. The polarisation of the Helium spin can be encoded by a phase quantity ϕ . We assume that the relationship between the generated field strength B and the current I flowing through the coil is linear, i.e. $B = B_{\text{eff}} \cdot I$ for some constant B_{eff} . If the solenoid has length L then the total accumulated phase has the form

$$\phi = \frac{\gamma}{V} B_{\text{eff}} I, \quad (4.13)$$

where V denotes the velocity of the He atom and γ is the gyromagnetic ratio of the He atom. Therefore if ϕ_i denotes the accumulated phase in the first coil and ϕ_f the phase in the second then

$$\phi = \phi_i + \phi_f = \gamma B_{\text{eff}} \left(\frac{I_i}{V_i} + \frac{I_f}{V_f} \right), \quad (4.14)$$

where we assume that the currents I_i, I_f and velocities V_i, V_f are different but the length L and constant B_{eff} is the same. We observe that the incoming velocity V_i is related to the incoming wavelength via (4.2), i.e. $V_i = p_i m^{-1} = h(m\lambda_i)^{-1}$, where λ_i denotes the wavelength of the beam. Consequently, (4.14) becomes

$$\phi = m\gamma B_{\text{eff}} h^{-1} (I_i \lambda_i + I_f \lambda_f). \quad (4.15)$$

Now suppose that polarisation is rotated in the xy -plane and is initially polarised in the x -direction. Then, assuming that the analyser near the detector is also setup to read the x -direction, the signal received has the form [58]

$$\begin{aligned} P_{x,I}(I_i, I_f) &= \langle \cos \phi \rangle_\rho \\ &= \int \rho(\lambda_i, \lambda_f) \cos \left(m\gamma B_{\text{eff}} h^{-1} (I_i \lambda_i + I_f \lambda_f) \right) d\lambda_i d\lambda_f. \end{aligned} \quad (4.16)$$

Here $\rho(\lambda_i, \lambda_f)$ denotes the *Wavelength Intensity Function* describing the distribution of helium atoms that reach the final polariser according to initial and final wavelengths. Equation (4.16) can be cleaned up by replacing I_i, I_f with new scaled variables:

$$\kappa_i = \frac{m\gamma B_{\text{eff}} I_i}{2\pi h}, \quad \kappa_f = \frac{m\gamma B_{\text{eff}} I_f}{2\pi h}. \quad (4.17)$$

Combining this with polarisation in the y -direction, (4.16) has the complex form

$$\begin{aligned} P(\boldsymbol{\kappa}) &= \int \rho(\boldsymbol{\lambda}) e^{2\pi i \boldsymbol{\kappa} \cdot \boldsymbol{\lambda}} d\boldsymbol{\lambda}, \\ \boldsymbol{\lambda} &= (\lambda_1, \lambda_2), \boldsymbol{\kappa} = (\kappa_1, \kappa_2) \in \mathbb{R}^2. \end{aligned} \quad (4.18)$$

A model example of a wavelength intensity function is given in Figure 4.5, along with possible phonon creation and annihilation events labelled. Another 2D example can be found in [64] for incoming/outgoing energy spectra. Assuming the features in the plot originate from phonon phenomena on the crystal surface, the classification into creation/annihilation/elastic originates from Equation (4.12):

$$E_f - E_i = \frac{\hbar^2}{2m} (k_f^2 - k_i^2) = \frac{\hbar^2}{2m} \left(\frac{1}{\lambda_f^2} - \frac{1}{\lambda_i^2} \right). \quad (4.19)$$

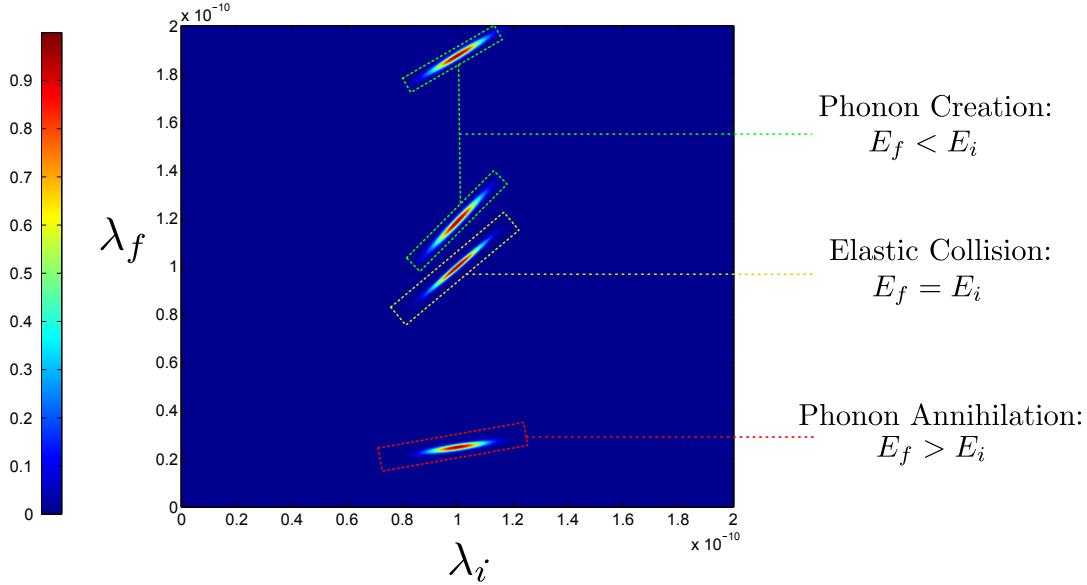


Figure 4.5: Example of a wavelength intensity function with key features highlighted. Classification of the features is deduced from Equation (4.19).

4.2.4 The Fourier Slice Theorem

Since the wavelength intensity function is supported around the average initial wavelength it could be viewed as a one-dimensional function that has been smoothed out into two-dimensions by the spread of initial wavelengths. Looking at Figure 4.5 as an example, we see that the key features of the wavelength intensity function can be broken down into lines of various slants which suggests that treating the function as one-dimensional would be advantageous.

It is precisely because of this decomposition into slanted regions that the two-dimensional problem is often reduced to a one-dimensional one using the Fourier slice theorem: Suppose we rotate the $\lambda = (\lambda_1, \lambda_2)$ coordinate system by an angle $-\alpha$ to a new system $\tau = (\tau_1, \tau_2) = R_\alpha(\lambda)$. Then one can derive the formula

$$\begin{aligned} P_\alpha(\kappa_1) &:= P(R_\alpha(\kappa_1, 0)) = P(\kappa_1 \cos \alpha, -\kappa_1 \sin \alpha) \\ &= \int \left(\int \rho(R_\alpha(\tau_1, \tau_2)) d\tau_2 \right) \exp(2\pi i \kappa_1 \tau_1) d\tau_1. \end{aligned} \quad (4.20)$$

Therefore if we let $\rho_\alpha(\tau_1)$ denote the integral of $\rho(\lambda)$ along the line $\{R_\alpha(\tau_1, \tau_2) : \tau_2 \in \mathbb{R}\}$ then we have

$$P(\kappa_1 \cos \alpha, -\kappa_1 \sin \alpha) = \int \rho_\alpha(\tau_1) \exp(2\pi i \kappa_1 \tau_1) d\tau_1. \quad (4.21)$$

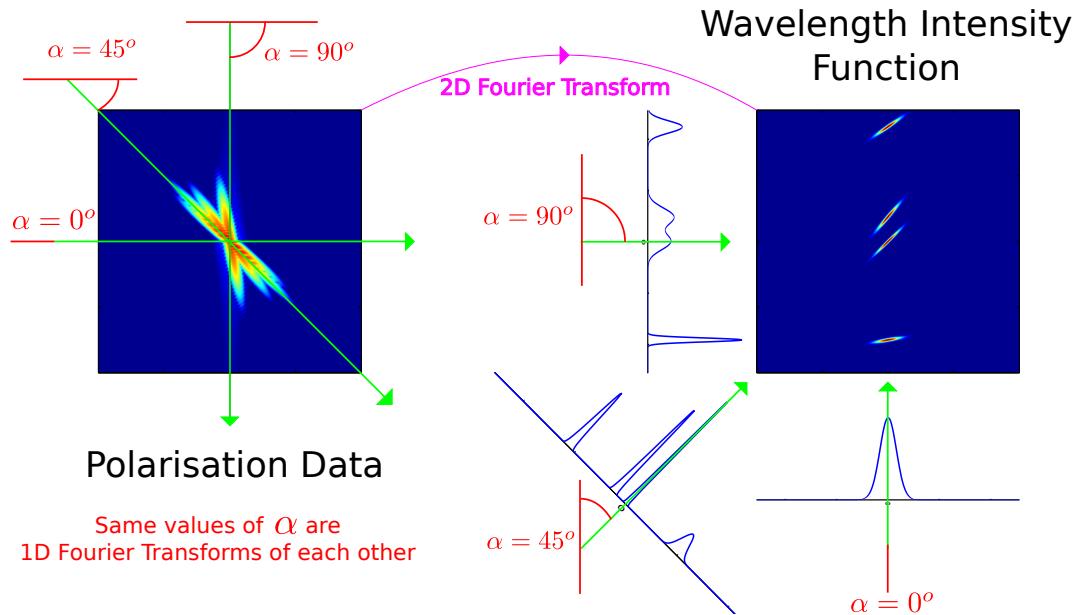


Figure 4.6: Demonstration of the Fourier slice phenomenon. On the left we have the polarisation data and on the right we have the (approximation to) the wavelength intensity function. On the right the green lines indicate the direction of integration and the corresponding projection is shown as graph in blue. The green lines on the left hand side represent the one-dimensional Fourier transforms of the projections shown on the right. The polarisation data intensity is shown on a log scale for the sake of readability.

This equation says that the restriction of P along the line $\{(\kappa \cos \alpha, -\kappa \sin \alpha), \kappa \in \mathbb{R}\}$, corresponds to the Fourier transform of ρ_α .

Figure 4.6 shows how the Fourier slice theorem applies to the wavelength intensity function shown in Figure 4.5. Notice that different angles of integration produce different results, especially when it comes to discerning the different slanted features. Since we know beforehand that an elastic peak lies along the line $\lambda_1 = \lambda_2$ we expect that an integration angle of $\alpha = \pi/4$ will produce the best results for resolving this feature as a single spike.

Since the spread of initial wavelengths is already rather concentrated in this example (and in practice), the Fourier slice theorem can be used to treat this distribution as a single point. This point, the average wavelength λ_{in} , is either known beforehand or deduced by projecting along the λ_2 -axis. With the incident wavelength $\lambda_1 = \lambda_{\text{in}}$ fixed, any projection can be converted to a function in terms of λ_2 only [58]:

$$\lambda_2 = \lambda_{\text{in}} \cot \alpha - \lambda_{\text{proj}} \sec \alpha, \quad (4.22)$$

where $\lambda_{\text{proj}} \in \mathbb{R}$ is a point on the line to which we project.

With this projection, we can effectively treat the problem (4.21) as a one-dimensional version of (4.18) with a new wavelength intensity function $\rho_{\lambda_1}(\lambda) := \rho_\alpha(\lambda)$

$$P_\alpha(\kappa) = \int_a^b \rho_{\lambda_1}(\lambda) e^{2\pi i \kappa \cdot \lambda} d\lambda, \quad \lambda, \kappa \in \mathbb{R}. \quad (4.23)$$

4.3 Applying CS to Phonon Detection

In this section we look at the theory and performance of CS on examples of phonon detection. We shall first look at its effects on one of the one-dimensional projections shown previously and then focus on real ${}^3\text{He}$ spectra where more exotic signal behaviour is present.

4.3.1 How this Work Developed

Before we begin going further it is worth clarifying what developments have been made by the collaboration in [3] and at what stage the SMF group was at before the collaboration began. What originally brought the two research groups together was previous work on identifying the benefits of applying CS to HeSE [65, 66] by the SMF group. This work was primarily for reconstructing the wavelength intensity function (namely the first stage in the full loop shown in figure 4.14 which is what we are focusing on in this section).

This means that the discrete approach in Section 4.3.2 below follows a similar approach to this previous work (although with different data sets). The application of continuous CS approach in section 4.3.4 is completely new to this field and is easier to integrate with the rest of the full loop as shall be shown later on in section 4.4.

4.3.2 Discrete CS Approach

In this section we shall assume that we have already reduced the problem to one-dimension and write $P_\alpha, \rho_{\lambda_1}$ from (4.23) as P, ρ . Now, since (4.23) is in the same form as (1.9) one can proceed in the same way to get to the truncated Fourier series approximation (1.11):

$$\begin{aligned} \rho(\lambda) &\approx \tilde{\rho}_N(\lambda) = \sum_{-N}^N \langle \rho, \exp_{k,\epsilon} \rangle \exp_{k,\epsilon}(\lambda), \quad \lambda \in [a, b], \\ \langle \rho, \exp_{k,\epsilon} \rangle &= \int_a^b \rho(\lambda) \exp_{k,\epsilon}(\lambda) d\lambda = \epsilon^{1/2} P(k\epsilon). \end{aligned} \quad (4.24)$$

One can discretise this approximation by breaking up the interval $[a, b]$ into a uniform grid of $2N + 1$ points

$$\lambda_{j,N} = a + \frac{b - a}{2N + 1} j, \quad j = 0, \dots, 2N, \quad (4.25)$$

leading to the following matrix equation:

$$f_j = \sum_{k=-N}^N A_{j,k} g_k, \quad j = 0, \dots, 2N, \quad f_j = \tilde{\rho}_N(\lambda_{j,N}), \quad (4.26)$$

where $g_k = \text{Constant}(k, \epsilon) \cdot P(k\epsilon)$ and $A = DFT$ is a FFT matrix.

Currently, to obtain the full vector $(f_j)_{j=0, \dots, 2N}$ we need to know the entire vector $(g_k)_{k=-N, \dots, N}$. If we only had knowledge of a fraction of the entries of f we can no longer use (4.26) to determine f directly as the problem is now underdetermined. Therefore, the problem is not well posed and has to be modified.

Noting that the inverse of the DFT matrix $B = A^{-1}$ is the conjugate transpose matrix ($B = A^\dagger$), which is also another DFT matrix (just with the frequencies flipped). Therefore, the matrix equation (4.26) can be inverted to give

$$g_j = \sum_{k=0}^{2N} B_{j,k} f_k, \quad j = -N, \dots, N. \quad (4.27)$$

Now suppose that $\Omega \subset \{-N, \dots, N\}$ denotes the set of indices corresponding to the samples of g that are measured and P_Ω denotes the projection onto these samples. With this notation $P_\Omega f$ denotes the vector of samples that are measured.

Therefore when we subsample from $\{-N, \dots, N\}$ equation (4.27) becomes

$$P_\Omega g = P_\Omega B f. \quad (4.28)$$

The classical CS approach is to solve this problem via the now well established l^1 recovery problem

$$\min_{z \in \mathbb{C}^N} \|Wf\|_{l^1} \quad \text{subject to} \quad P_\Omega g = P_\Omega B f, \quad (4.29)$$

where W is some transformation that should make g sparse. Typically this is a wavelet transformation. We can then solve this kind of problem quickly and conveniently using convex solvers such as the (SPGL1) algorithm [12, 13].

The classical idea of CS is that Ω should be chosen uniformly at random. Recall

from (1.7) that in this case the number of samples $m = |\Omega|$ should satisfy

$$m \gtrsim N \cdot \mu(BW^{-1}) \cdot s \cdot \log(N), \quad (4.30)$$

in order to guarantee successful recovery with high probability. In the case where $U = BW^{-1}$ as above with any wavelet transform W we have that $\mu(U) = 1$. In this case uniform random sampling gives poor reconstruction results and one should instead use variable density sampling we now describe.

4.3.3 Quick Description of Multilevel Sampling

As demonstrated in Section 1.6.2, the correct sampling approach will depend on the actual structure of the signal, and given the popular use of wavelets it is natural to analyse the structure these systems yield. In particular, the signal coefficients have a very specific level structure when decomposed into such bases. This is known as sparsity in levels.

Sparsity in levels: Let x be a \mathbb{C}^N vector. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \dots < M_r$ and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \dots, r$, where $M_0 = 0$. We say that x is (\mathbf{s}, \mathbf{M}) -sparse if, for each $k = 1, \dots, r$, $\Delta_k := \text{supp}(x) \cap \{M_{k-1} + 1, \dots, M_k\}$, satisfies $|\Delta_k| \leq s_k$. This known structure can be utilised when designing the sampling strategy and is the motivation behind multilevel sampling.

Multilevel sampling: Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, and suppose that $\Omega_k \subseteq \{N_{k-1} + 1, \dots, N_k\}$, $|\Omega_k| = m_k$, $k = 1, \dots, r$, are chosen uniformly at random, where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N}, \mathbf{m}} = \Omega_1 \cup \dots \cup \Omega_r$ as an (\mathbf{N}, \mathbf{m}) -multilevel sampling scheme. The key is that in the case of Fourier sampling, represented by the B above, combined with a wavelet transform W such that the recovery problem becomes (4.29), the multilevel sampling should match the level structure of the wavelets. More precisely, $\mathbf{N} = \mathbf{M}$. In this case, if x is (\mathbf{s}, \mathbf{M}) -sparse with total sparsity $s = s_1 + \dots + s_r$ and $s_1 = M_1 = m_1$. Then the total number of samples needed is

$$m = m_1 + \dots + m_r \gtrsim (s_1 + s_2 + s_3 + \dots + s_r) \log(N). \quad (4.31)$$

For a more in-depth analysis and explanation see Sections 1.4, 2.8 or [32].

4.3.4 Continuous CS for Phonon Spectra

Observe that the Fourier series approximation $\tilde{\rho}_N$ (see (4.24)) could have been evaluated at any point in the interval $[a, b]$, however, the best we can achieve using the FFT/discrete wavelet approach is to reconstruct $\tilde{\rho}_N$ at the values $\lambda_{j,N}$, even though we are still working with the same set of Fourier data. Do we really have to pay this price in order to be able to compress this problem?

Motivated by the benefits of (4.24), one can try approximating the wavelength intensity function ρ in terms of a new *Reconstruction Basis* σ_n , $n \in \mathbb{N}$:

$$\tilde{\rho}(\lambda) \approx \sum_{n=1}^N \langle \rho, \sigma_n \rangle \sigma_n(\lambda), \quad \lambda \in [a, b]. \quad (4.32)$$

For technical reasons, one often requires these functions to form an orthonormal basis of $L^2[a, b]$, although this condition can be relaxed to other groups of functions like frames [67]. For this section we shall be using Daubechies wavelets exclusively as our reconstruction basis, which we cover in Section 4.3.5. Let's quickly discuss *why* one would want to work with another basis.

Apart from the benefit of the keeping the problem continuous, one also has the freedom to *choose* which basis σ_n to work with, making the approach more versatile than a straight DFT approach, where we are essentially forced to work with a pixel basis every time.

Furthermore, the notion of *sparsity* is now in terms of the coefficients $\langle \rho, \sigma_n \rangle$, which means we have the additional advantage of choosing a basis that makes the function ρ sparse. As we shall see, this opens up the possibility of using compressed sensing where traditional sparsity does not hold. In addition, this approach is closer to the philosophy that ρ being sparse should relate to ρ having low information content; a choice of basis σ_n that makes ρ sparse tells us how to (approximately) express the function ρ with a few non-zero coefficients.

Since one is still sampling data that corresponds to Fourier coefficients of ρ , it is impossible to exclusively work with their choice of basis σ_n . Instead one has to convert Fourier series coefficients into coefficients in the basis σ_n . This is achieved by working with the change of basis matrix for the two bases:

$$B_{k,n} = \langle \sigma_n, \exp_{k,\epsilon} \rangle, \quad n \in \mathbb{N}, \quad k \in \mathbb{Z}. \quad (4.33)$$

We can then mimic (4.29) by solving a convex optimisation problem:

$$\begin{aligned} \operatorname{argmin}_{h \in \ell^2(\mathbb{N})} & \left\{ \|h\|_1 = \sum_{i \in \mathbb{N}} |h_i| : P_\Omega B h = P_\Omega f \right\}, \\ \ell^2(\mathbb{N}) &= \left\{ h \in \mathbb{C}^\mathbb{N} : \sum_{k=1}^{+\infty} |h_k|^2 < \infty \right\}. \end{aligned} \quad (4.34)$$

Again P_Ω denotes the projection onto the samples we've taken. In practice we cannot solve for the infinite solution to (4.34), therefore we truncate the reconstruction basis in a similar fashion to how we truncate the Fourier basis. This means we end up solving

$$\operatorname{argmin}_{h \in \mathbb{C}^N} \left\{ \|h\|_1 : P_\Omega B P_N h = P_\Omega f \right\}, \quad (4.35)$$

where P_N denotes the projection onto the first N functions in the reconstruction basis. This problem is now numerically feasible since the submatrix $P_\Omega B P_N$ is now finite. The solution to (4.35), let's say h^* , is recognised as the (approximate) wavelet coefficients of the intensity function. We can then use these wavelet coefficients to compute an approximation to ρ evaluated at any point on the interval $[a, b]$ by following (4.32):

$$\rho(\lambda) \approx \sum_{n=1}^N h_n^* \sigma_n(\lambda), \quad \lambda \in [a, b]. \quad (4.36)$$

From here one can use the same multilevel sampling techniques as the discrete case to reconstruct (\mathbf{s}, \mathbf{M}) -sparse coefficients. Moreover, the sampling rule (4.31) also applies in this case. These details can be found in Section 1.4.

4.3.5 Wavelet Bases

The family of reconstruction bases that we shall be using for HeSE in this thesis is the *Daubechies wavelets* [21, 22]. A wavelet basis is usually constructed from something called a *scaling function*. This is somewhat similar to how pixels are used to make up screen resolutions on computer monitors. The simplest example of a scaling function is the Haar scaling function which is essentially one pixel of unit length:

$$\phi : \mathbb{R} \rightarrow \mathbb{R}, \quad \phi(x) = \begin{cases} 1, & x \in [0, 1], \\ 0, & \text{otherwise.} \end{cases} \quad (4.37)$$

We can then use this function to create different resolution levels (one for each value of j) by scaling and translating this function appropriately:

$$\phi_{j,k}(x) = 2^{-j/2}\phi(2^j x - k), \quad j, k \in \mathbb{Z}, \quad V_j = \overline{\text{span}}\{\phi_{j,k} : k \in \mathbb{Z}\}, \quad (4.38)$$

where “span” means the linear span of the functions treated as vectors (in $L^2(\mathbb{R})$). Notice that if we fix the resolution level j then the functions $\phi_{j,k}$ for $k \in \mathbb{Z}$ are orthogonal to each other but functions from different resolution levels are not necessarily orthogonal. This is where the *mother wavelet* ψ comes in. The mother wavelet is a linear combination of the $\phi_{1,k}$ ’s; in the case of the Haar scaling function the corresponding Haar wavelet is defined as

$$\psi(x) = \phi_{1,1}(x) - \phi_{1,0}(x) = \begin{cases} -1, & x \in [0, 1/2), \\ 1, & x \in [1/2, 1), \\ 0, & \text{otherwise.} \end{cases} \quad (4.39)$$

The wavelet ψ is also scaled and shifted like ϕ :

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}, \quad W_j = \overline{\text{span}}\{\psi_{j,k} : k \in \mathbb{Z}\}. \quad (4.40)$$

The shifts of ψ ($= W_0$) and ϕ ($= V_0$) generate V_1 and moreover these functions are all orthogonal with each other. This means we can use both the $\phi_{j,k}$ ’s and $\psi_{j,k}$ ’s to construct orthonormal bases that cross several resolution levels. These functions even have the same convergence properties as the Fourier basis, namely that

$$\rho(x) = \sum_{k \in \mathbb{Z}} \left(\langle \rho, \phi_{J,k} \rangle \phi_{J,k}(x) + \sum_{j=J}^{+\infty} \langle \rho, \psi_{j,k} \rangle \psi_{j,k}(x) \right). \quad (4.41)$$

In practice we work over the interval and not over the whole real line. By scaling the domains of the functions appropriately we can assume this interval is $[0, 1]$. For a function $f \in L^2(\mathbb{R})$ let $\tilde{f}(x) = \sum_{k \in \mathbb{Z}} f(x - k)$ denote the $[0, 1]$ -periodised version² of f . Then the periodised version of (2.2) becomes (for $J \geq 0$)

$$\tilde{\rho}(x) = \sum_{k=1}^{2^J} \langle \tilde{\rho}, \tilde{\phi}_{J,k} \rangle \tilde{\phi}_{J,k}(x) + \sum_{j=J}^{+\infty} \sum_{k=1}^{2^j} \langle \tilde{\rho}, \tilde{\psi}_{j,k} \rangle \tilde{\psi}_{j,k}(x). \quad (4.42)$$

²We assume that the function is either compactly supported or decays sufficiently fast for this periodisation to be well-defined.

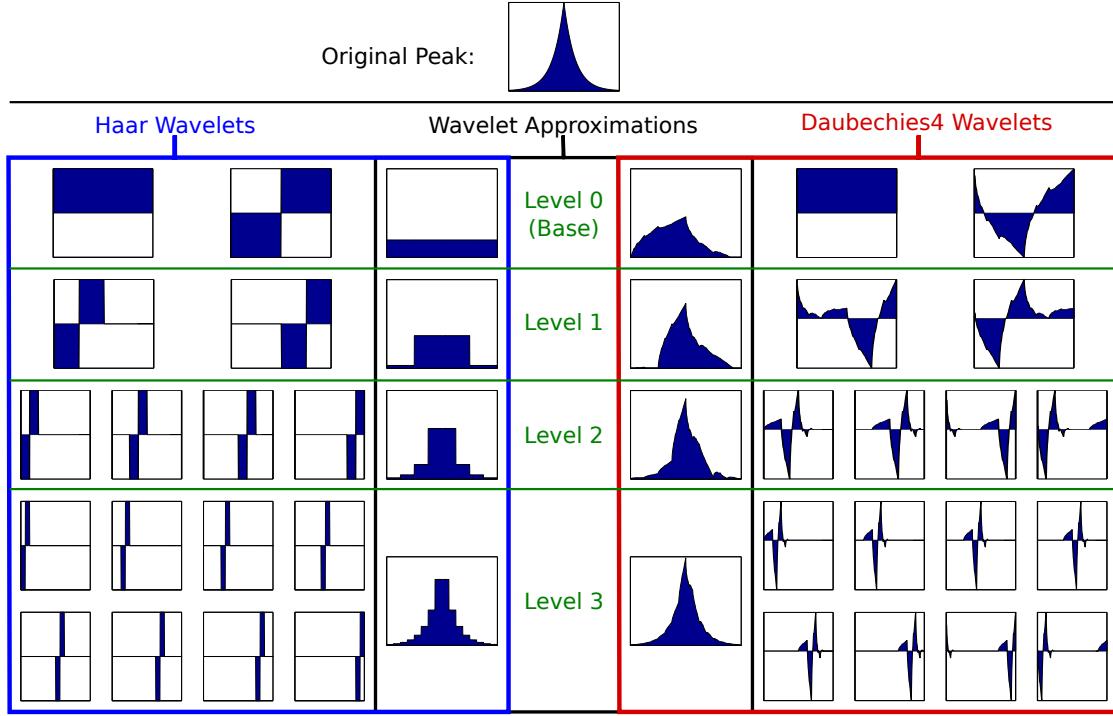


Figure 4.7: This figure depicts two periodised wavelet bases (Haar and Daubechies4) along with an example of how they are used to construct a simple exponential peak. Scales are not shown. For the purposes of approximation, each wavelet level includes all previously levels listed above it. Since we are working with *periodic* wavelets, the Daubechies4 scaling function is *not* present here.

Consequently we deduce that the following collection of functions is a basis for $L^2[0, 1]$:

$$B_w = \{\tilde{\phi}_{J,k} : k = 1, \dots, 2^J\} \cup \bigcup_{j=J}^{\infty} \{\tilde{\psi}_{j,k} : k = 1, \dots, 2^j\}. \quad (4.43)$$

Two examples of periodic wavelet bases are shown in Figure 4.7 along with an application to a single exponential peak.

Apart from the Haar basis, we have yet to talk about other Daubechies wavelet bases³. We shall not go into a full description here but their beneficial properties include:

1. Resolution Levels: By working over multiple resolution levels we now have some control over the structure of our reconstruction. In particular, we can vary this structure to improve the sparsity of a given problem by changing the starting level J .
2. Vanishing Moments: They are particularly useful for approximating functions with

³We are using the convention that a Haar wavelet is a Daubechies2 wavelet here.

certain degrees of smoothness. This is because a Daubechies $2n$ wavelet is orthogonal to all polynomials of degree $n - 1$ or less and a smooth function (say f) tends to have a fast decaying Taylor expansion. This means that when we calculate $\langle f, \psi_{j,k} \rangle$ the wavelet will eliminate the first n terms in the Taylor expansion leaving a (hopefully) small remainder.

3. Minimal Support: Daubechies wavelets have the unique property that they have minimal support size for a given number of vanishing moments. This means that they are highly localised.

4.3.6 Simulated 1D Example

For consistency with previous sections we shall first work with the 45 degree projection shown earlier. Although this is a simplified model it clearly demonstrates some of the basic properties of compressed sensing. Reconstructions are shown in Figure 4.8.

Recall that we project along a 45 degree angle in an attempt to reduce the spread caused by the inaccuracy of the wavelength of the incident ${}^3\text{He}$ beam. In particular, since we know that there will always be an elastic feature in the wavelength intensity function which itself is slanted at a 45 angle, this choice is seen as ideal for refocusing the various phonon features to be closer to that of a delta spike. Not only is this useful in preventing features from overlapping each other but this also increases sparsity which is ideal for compressed sensing; if features are sparse then by the rule (4.30) we can subsample to a great degree since $s(f)$ is small.

If one goes for a uniformly random approach to subsampling (as in Reconstruction A) then there is only so far that one can go before problems occur. At around 20% the reconstruction becomes unreliable in recovering the least sparse of the features on the far right. Around 30% however, the rightmost features is typically reconstructed but it is nonetheless unreliable. Compressed Sensing, at the very least in theory, depends on random sampling and therefore its performance also has some degree of randomness. Fortunately one finds that both in theory and in practice if one samples significantly above the threshold given by estimates of the form (4.30) then the probability of failure is exponentially small thanks to guarantees like (1.7).

There is however an even more effective way of reliably reconstructing the rightmost feature which involves sampling more of the lower frequency coefficients while throwing out the larger ones (as in Reconstruction B). This can be seen as a cross between traditional CS and classical Fourier series approximation which we shall see again in the

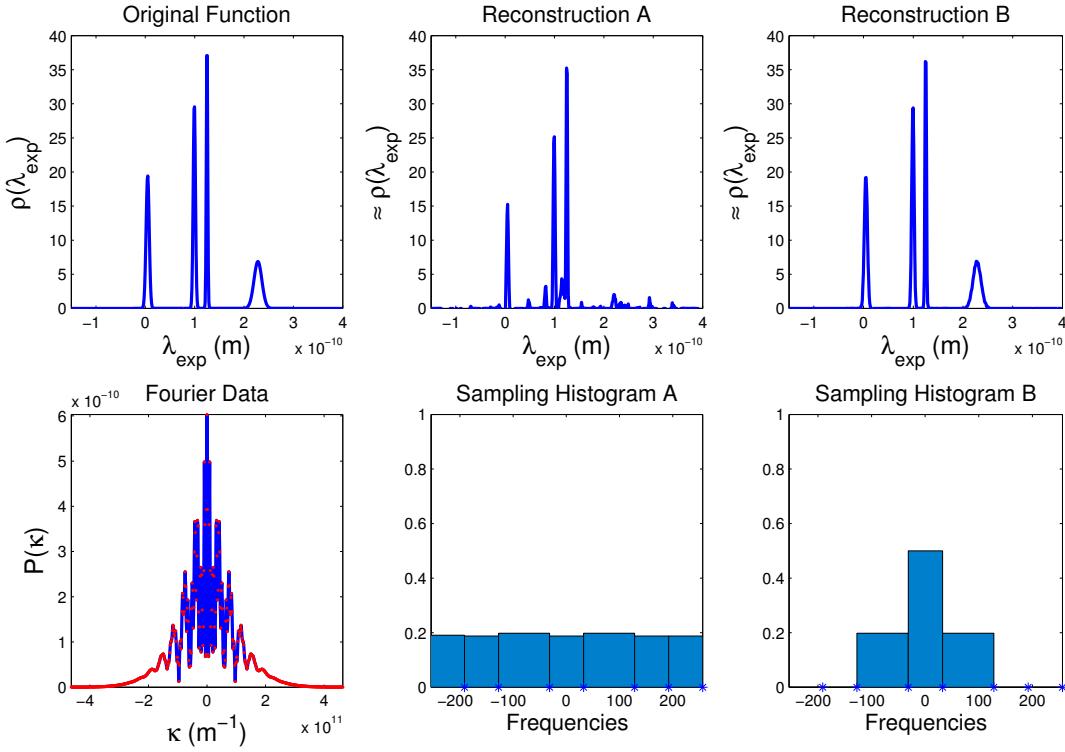


Figure 4.8: Demonstration of conventional CS for the 1D 45 degree projection shown in Figure 4.5. Samples are taken from the Fourier data according to the sampling histograms shown. As the method involves random sampling the outcome (and therefore accuracy) of each reconstruction varies. Sampling pattern A is unreliable in reconstructing the rightmost feature as it is the least sparse of the four peaks while sampling pattern B remedies this by taking more of the lower frequency values that it depends upon. The frequencies listed on the histograms are relative to the fundamental frequency of the reconstruction domain and not their true (non-integer) values. Reconstructions are at a resolution of 512 data points.

next example.

4.3.7 Real Phonon Spectrum

As we have already mentioned, real phonon spectra contain more exotic features than the simulation given in the previous example. Naturally noise adds to the data due to some experimental uncertainties in the apparatus, but more unusual are the relative sizes and shapes of the various features.

In Figure 4.9 we have uniform sampling reconstructions for a typical Gold spectra with projection at 45 degrees to focus on the elastic peak, which is the only clearly visible feature in the graphs. The peak is extremely fine and is in fact even smaller than the pixel resolution used for reconstruction (2048) which can be determined from the observation that the Fourier data has yet to decay to zero near the highest frequencies.

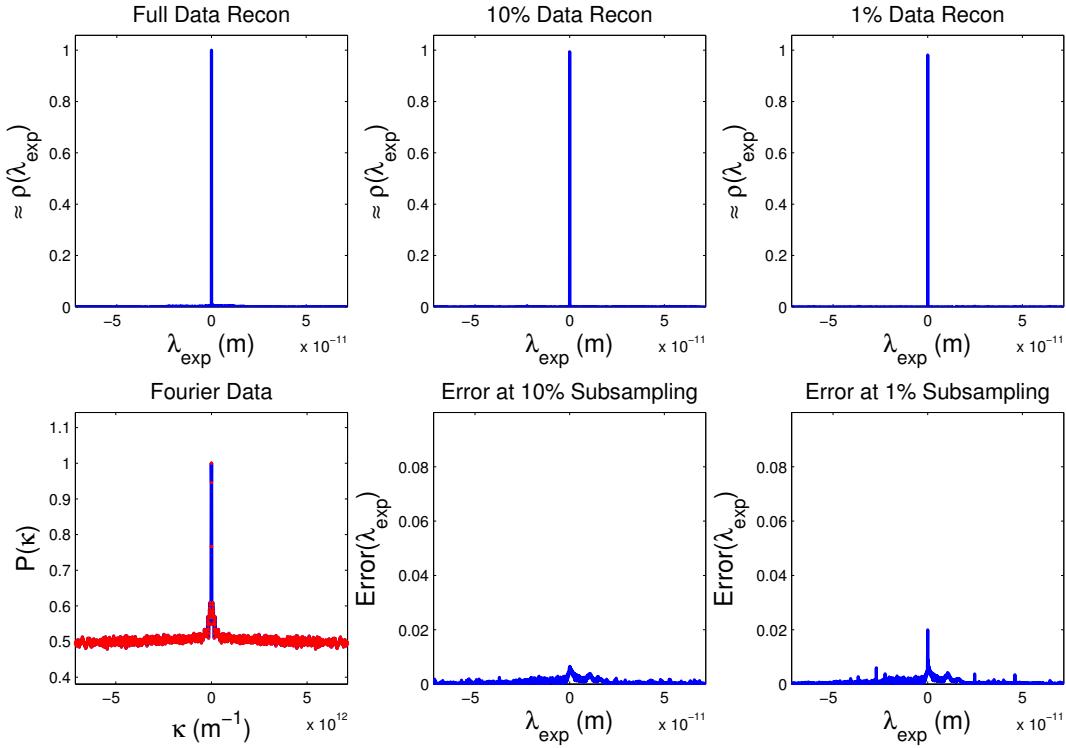


Figure 4.9: Compressed sensing reconstructions for a gold phonon spectrum. With this choice of viewing range, only the elastic peak can be clearly seen. For a zoomed in closeup on other features see Figure 4.10. In this Figure the sampling is performed uniformly at random which is ideal for reconstructing the highly sparse central peak. Frequencies sampled are from the range $\{-1024, \dots, 1023\}$ and reconstructions are at a resolution of 2048.

Consequently this is an ideal situation for basic sensing since this feature is almost as sparse as can be. Hence, one can subsample to a much greater degree (e.g. 1%) than in the previous example.

However, what has happened to the non-elastic features in this spectrum? At first one might come to the conclusion that they are not there at all, however, focusing on a small part of this spectrum reveals features that are over 200 times smaller than the measured intensity of the large elastic spike. Figure 4.10 shows various CS reconstructions zoomed in on this region. Notice that we still have the elastic peak visible, along with a couple of more (and less sparse) features. Like in the previous example, we expect the smooth features to be more dependent upon the lower frequency samples. Therefore, when we attempt to take just the first 10% of samples all from the lowest frequencies (the *linear* reconstruction) we find that these features are at the very least present, unlike the 10% uniform sampling approach where only the central peak remains. On the other hand, the central peak suffers from Gibbs artefacts which manifests themselves as wave-like

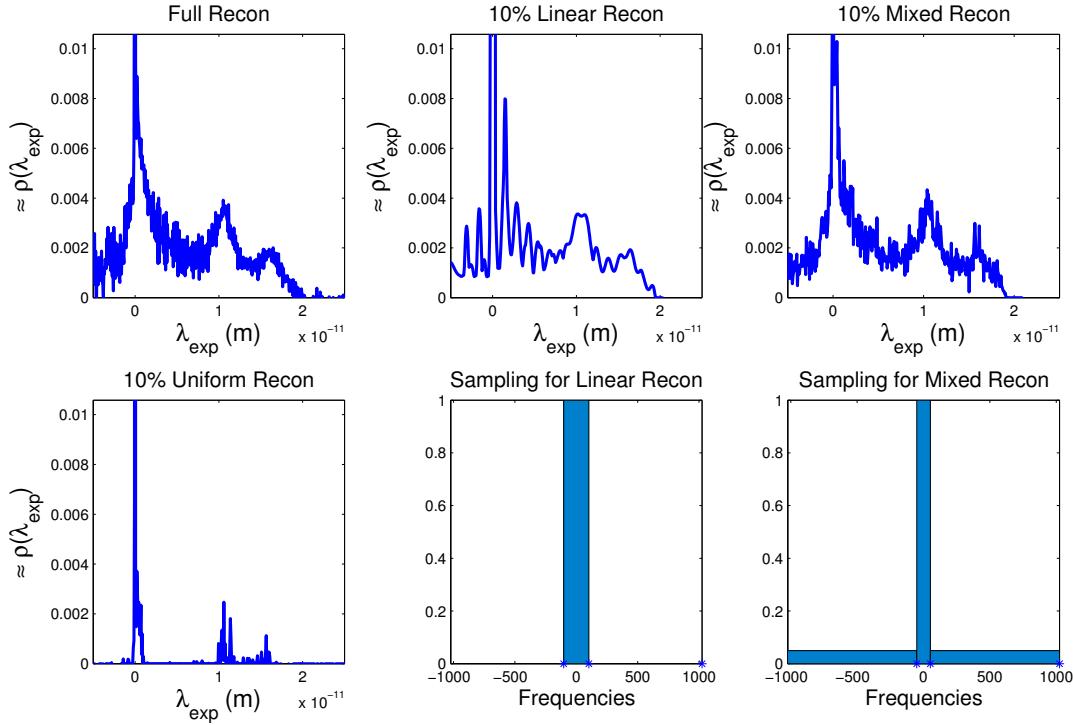


Figure 4.10: Compressed sensing reconstructions for the same gold phonon spectrum, zoomed in so that features beside the elastic peak are visible. For a complete picture of the spectrum consult Figure 4.9. Notice that the smaller features shown here are over 200 times smaller than the elastic peak. Reconstructions shown here are not only uniform (as in the bottom left graph) but also linear (i.e. straight Fourier series) and non-linear examples using roughly the same number of samples across each.

features near the central peak as well as broadening of the peak itself.

Instead, one can opt for a mix of these methods by taking the first 5% of samples from the lowest frequencies and the other 5% taken uniformly from the rest. This approach empirically performs the best out of the three methods, resolving the low resolution features without the Gibbs artefacts of the linear approach.

4.3.8 Comparing CS Techniques

In this section we look at an example of how the continuous wavelet approach to compressed sensing can be used to tackle problems that are beyond the capabilities of the traditional compressed sensing approach described earlier.

Prior to data acquisition the spacing in current (equivalent to spacing in κ) must be chosen, which in turn determines the length of the wavelength window $[a, b]$ that the wavelength intensity function ρ is constructed over. If ρ is not truly supported on this window we instead reconstruct the periodised version of ρ . In particular, if peaks in the

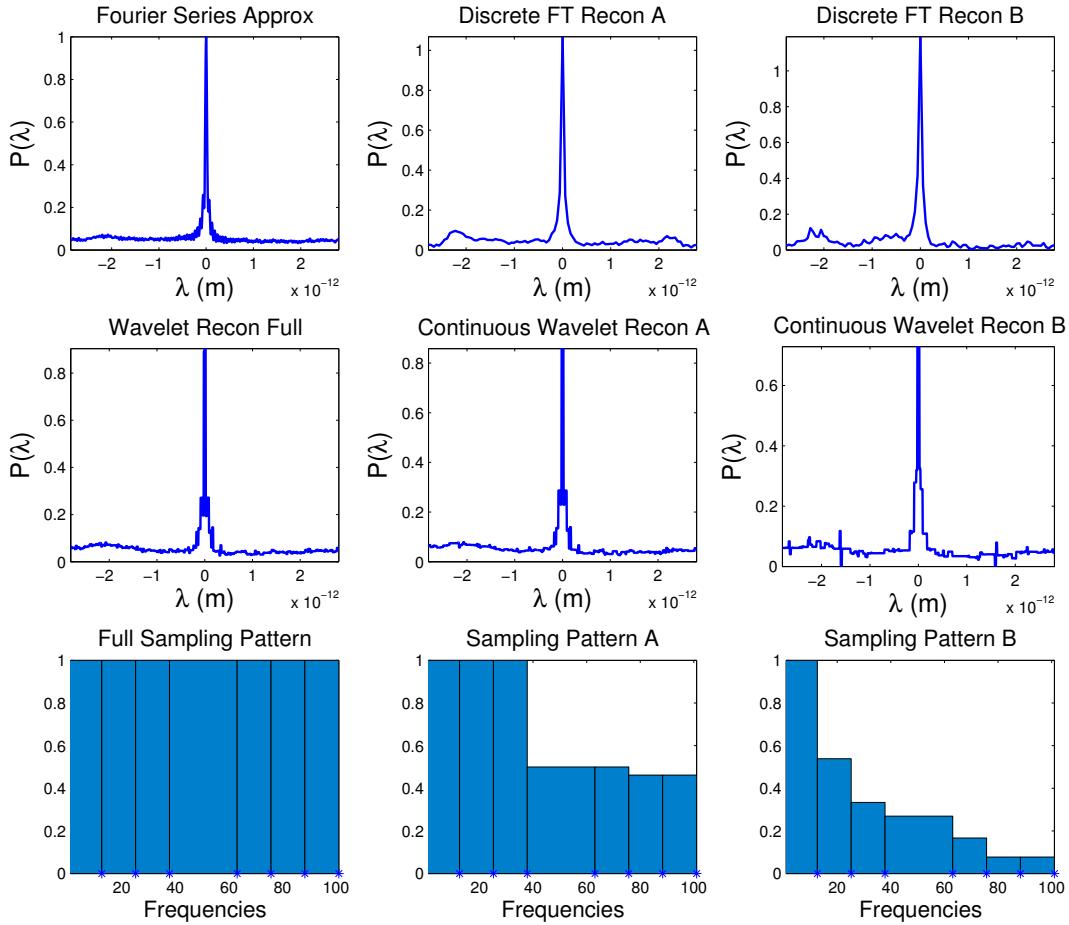


Figure 4.11: Reconstruction of a CoPc Quasi-Elastic Spectra with a noticeable baseline feature. In the the DFT reconstructions the baseline level of around 0.1 is no longer flat leaving bumpy artefacts while the wavelet reconstructions preserve this flat feature. As we are only subsampling from 101 frequency points, considerable Gibbs artefacts are present in the Fourier series approximation. The DFT reconstructions have a resolution of 101 points, while the continuous Fourier series and wavelet reconstructions have been rasterised at a resolution ten times this number.

intensity function decay particularly slowly relevant to the window then the intensity function will stay considerably above zero throughout that window. Because of the this, the traditional compressed sensing approach applied earlier cannot be used successfully here as the function is maximally non-sparse.

However, if one recalls the wavelet reconstruction bases that are used, then one quickly notices that they both have a constant function as the first basis function. This effectively means that the base level caused by slow decay is captured by this single basis function, which keeps the function sparse in these bases.

Figure 4.11 compares the two compression techniques for cobalt phthalocyanine (CoPc , $\text{C}_{32}\text{H}_{16}\text{CoN}_8$) on $\text{Ag}(001)$. This spectra has an observable baseline feature. The

full set of polarisation data points only corresponds to the first 101 frequencies and therefore there is noticeable Gibbs artefacts around the elastic peak in the Fourier series approximation. This strongly suggests that the Fourier series approximation here is not a particularly accurate approximation to the true intensity function.

Furthermore, the wavelet approximation aims to reconstruct the true underlying continuous wavelength intensity function, unlike the DFT approach which attempts to reconstruct a discretised form of the Fourier series approximation. Consequently, even with full sampling, the wavelet reconstruction is noticeably different to the Fourier series approximation. This reflects the fact that, as we are handling real data, we cannot directly compare to the true underlying wavelength intensity function.

Regardless we clearly observe that the baseline feature is preserved under subsampling using wavelets where the DFT approach clearly fails, matching predictions based on sparsity observations earlier. While we are able to subsample to a reasonable degree here ($\approx 33\%$), one should ideally work with a larger range of frequencies to truly exploit the benefits of this approach, i.e. subsampling from polarisation data with thousands of points rather than hundreds.

4.4 Background of Helium Spin-Echo Part 2 - Molecular Diffusion

The focus of this section is on studying diffusive properties of surfaces. As mentioned earlier, if one wants to determine how effective a surface behaves, e.g. for catalysis, then one must study how molecules move on top of the surface over time. This is very different to the previous phonon examples covered earlier because we are no longer just considering the motions of nuclei in the lattice. Instead we have species (molecules, atoms) diffusing on top of the surface that are interacting with each other.

Three models of diffusion are given in Figure 4.12. An important question is how one can differentiate these three types of diffusive regimes on a surface by using HeSE.

4.4.1 Scattering Cross Sections and the Van Hove Formalism

The van Hove formalism was initially developed for thermal neutron scattering [48] and we shall mention a few key results from this theory. This theory establishes a relationship between the neutron spectra (as a function of energy and momentum transfer) and

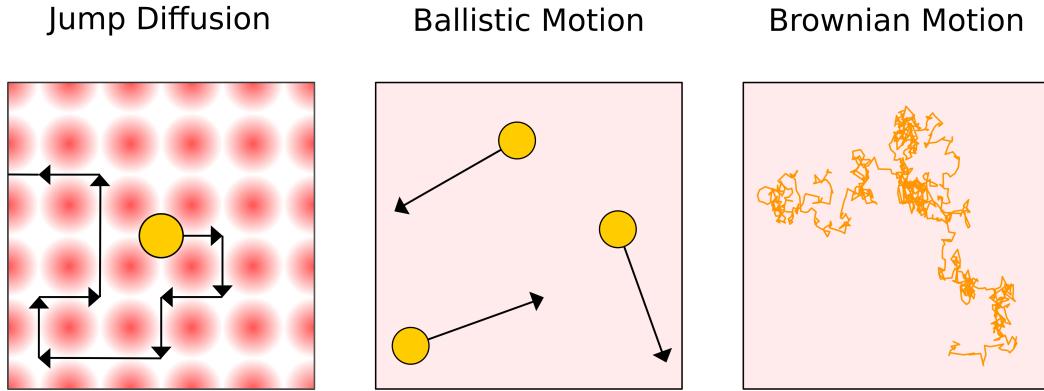


Figure 4.12: A depiction of three diffusive regimes. Orange denotes molecules on the surface and red denotes the surface potential. In the jump diffusion case the molecules are assumed to move instantly between vacant sites where the potential energy is smallest.

the dynamics of the nuclei within the sample. This theoretical approach makes some assumptions that generally holds well for neutrons, for example

- The incoming beam of neutrons has a fixed incident wavevector \mathbf{k} (and therefore a fixed incident energy). This is equivalent to monochromatisation of the beam.
- The potential for scattering of each nuclei is modelled as a Fermi pseudopotential, which is a delta spike at its position. This does *not* hold well for ${}^3\text{He}$.
- A scattering neutron interacts with the bulk potential at most once. This is carried through to ${}^3\text{He}$ scattering (multiple scattering events can be considered [68]).
- Fermi's golden rule is applicable to the combined neutron/scattering system.

With these assumptions one can compute the *Scattering Cross Section* $\frac{d^2\sigma}{d\Omega dE}$, defined as the number of neutrons scattered through a solid angle $d\Omega$ with change of energy dE divided by the flux of incident neutrons.

After some calculations the following formula for the cross section can be derived [48, Eq. 4.13]:

$$\left(\frac{d^2\sigma}{d\Omega dE} \right)_{\text{inc}} = \frac{\sigma_{\text{inc}}}{4\pi} \frac{k'}{k} N S_i(\mathbf{Q}, \omega), \quad (4.44)$$

where σ_{inc} is the incoherent⁴ scattering constant, $k = |\mathbf{k}|$, $k' = |\mathbf{k}'|$ are the magnitudes of the incoming and outgoing scattering wavevectors, $\mathbf{Q} = \mathbf{k}' - \mathbf{k}$, N is the total number of

⁴The scattering cross section is typically broken in coherent and incoherent terms, which correspond to self and distinct diffusion respectively. For convenience we work with the incoherent term only.

nuclei, ω denotes the change in energy (from (4.3)) and $S_i(\mathbf{Q}, \omega)$ is called the *Scattering Function* (SF). We shall drop the small i from here onwards.

What makes (4.44) particularly useful is that the structure factor is (by definition) related to the motion of nuclei in the bulk via multiple Fourier transforms:

$$\begin{aligned} G(\mathbf{R}, t) &= \int_{\mathbb{R}} \exp(-2\pi i \mathbf{Q} \cdot \mathbf{R}) I(\mathbf{Q}, t) dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^3} \exp(-2\pi i (\mathbf{Q} \cdot \mathbf{R} + \omega t)) S(\mathbf{Q}, t) d\mathbf{Q} dt. \end{aligned} \quad (4.45)$$

The function $G(\mathbf{R}, t)$ is often called the *van Hove correlation function* and $I(\mathbf{Q}, t)$ the *Intermediate Scattering Function* (ISF). The correlation function $G(\mathbf{R}, t)$ can be interpreted as the probability that a particle starting at the origin at time 0 will be at position \mathbf{R} at time t . By combining this fact with the scattering cross section formula (4.44), we see that the cross section contains probabilistic information regarding the dynamics of the nuclei in the bulk.

4.4.2 Extension to HeSE

Formula (4.44) can be adapted [58] to the surface sensitive HeSE approach using

$$\frac{d^2\sigma}{d\Omega dE} = S_i(\mathbf{K}, \omega) \cdot |F(\mathbf{K}, \omega)|^2, \quad \mathbf{K} \in \mathbb{R}^2, \omega \in \mathbb{R}, \quad (4.46)$$

where \mathbf{K} now denotes $\mathbf{k}_f - \mathbf{k}_i$ projected onto the surface (like in (4.11)) and $F(\mathbf{K}, \omega)$ denotes the *Form Factor* that takes into account the fact that ${}^3\text{He}$ atoms are scattered from a surface potential. This contrasts with the nuclear scattering of neutrons. For simplicity, we shall be assuming that the form factor can be compensated for when interpreting the scattering data, which effectively means setting this term to be equal to 1. Similarly (4.45) still holds but for the two-dimensional equivalents of $G(\mathbf{R}, t)$, $I(\mathbf{Q}, t)$ and $S(\mathbf{Q}, t)$ (with $\mathbf{Q} \in \mathbb{R}^3$ replaced by $\mathbf{K} \in \mathbb{R}^2$ and now $\mathbf{R} \in \mathbb{R}^2$)

As we shall see shortly in the next section, one often does not have the luxury of knowing the scattering function $S(\mathbf{K}, \omega)$ on the entire (\mathbf{K}, ω) space. This means that one cannot easily calculate the van Hove correlation function directly. Instead, it is sometimes preferable to infer properties of $G(\mathbf{R}, t)$ from partial knowledge of $S(\mathbf{K}, \omega)$. One such example is the *Half Width Half Maximum* (HWHM), which is the half life of $S(\mathbf{K}, \omega)$ for the variable ω as a function of \mathbf{K} , namely

$$\Gamma(\mathbf{K}) = \inf\{\omega > 0 : |S(\mathbf{K}, \omega)| = |S(\mathbf{K}, 0)|/2\}. \quad (4.47)$$

Different types of diffusion are presented in Figure 4.12. For the Brownian/Ballistic/Jump Motion the HWHM shows a quadratic/linear/periodic dependence upon \mathbf{K} [58].

Similarly, one can work with the ISF to try and classify the dynamical behaviour. The *Dephasing Rate* $\alpha(\mathbf{K})$, which describes the decay in t of the ISF as a function of \mathbf{K} is an example of this. Formally,

$$\alpha(\mathbf{K}) = \inf\{t > 0 : |I(\mathbf{K}, t)| = e^{-1}|I(\mathbf{K}, 0)|\}. \quad (4.48)$$

Just like the HWHM, for the Brownian/Ballistic/Jump Motion the dephasing rate shows a quadratic/linear/periodic dependence upon \mathbf{K} .

4.4.3 The Full Loop - From Wavelength to Energy

Now we connect up the dots between the van Hove formalism described in Section 4.4.1 and the experimental background given in Section 4.2.3.

Suppose we have completed the approach used to detect phonons, where we have used equation (4.18) to reconstruct a wavelength intensity function $\rho(\lambda_i, \lambda_f)$ where the initial and final ${}^3\text{He}$ wavelengths are λ_i and λ_f (we shall worry about the alternative projection approach later). The important question we are now faced with is: how does this intensity function relate to the scattering cross section?

Recall that by (4.46) we know that the scattering cross section is expressed in terms of change of surface momentum \mathbf{K} and energy $E = \hbar\omega$. Therefore if one wants to convert wavelength intensity to intensity in terms of energy or wavelength one first needs to change variables (recalling (4.11) and (4.12)):

$$\begin{aligned} \omega &= \frac{\Delta E}{\hbar} = \frac{\hbar}{2m} k_f^2 - \frac{\hbar}{2m} k_i^2 = \frac{\pi\hbar}{m} \left(\frac{1}{\lambda_f^2} - \frac{1}{\lambda_i^2} \right) \\ K &= k_f \sin \theta_i - k_i \sin \theta_f = \frac{2\pi}{\lambda_f} \sin \theta_i - \frac{2\pi}{\lambda_i} \sin(\gamma - \theta_i). \end{aligned} \quad (4.49)$$

Here $\gamma = \theta_i + \theta_f$ is the total scattering angle between the source/surface/detector setup which cannot be changed for the Cambridge spin-echo apparatus. Instead one tilts the surface in order to vary the incident angle θ_i , which in turn determines θ_f . The actual direction of the vector \mathbf{K} is determined by the geometry of the apparatus (formally when we restrict \mathbf{K} to the plane in the source/surface/detector setup it becomes a scalar hence why we only have a scalar K in (4.49)).

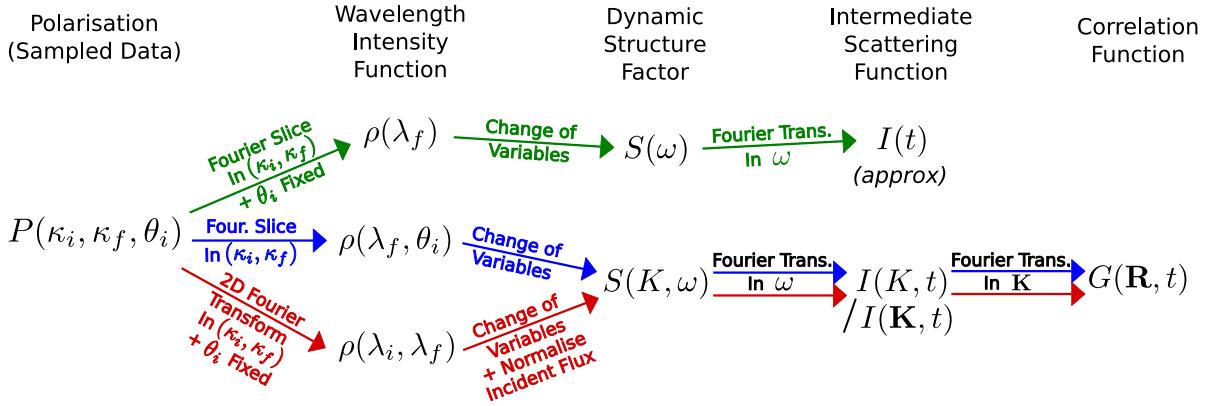


Figure 4.13: An outline of how one can work from polarisation data through to the ISF and the van Hove Correlation function. The red, blue and green colours correspond to three methods described in Section 4.4.3.

From (4.49) we see that the initial scattering angle θ_i is also an important variable in our experiments, therefore it is convenient to explicitly declare this dependency by rewriting $\rho(\lambda_i, \lambda_f)$ as $\rho(\lambda_i, \lambda_f, \theta_i)$. Our goal is to convert knowledge of $\rho(\lambda_i, \lambda_f, \theta_i)$ to knowledge of the scattering function and then to the intermediate scattering function. We shall be discussing three methods of achieving this:

1. Keeping θ_i fixed but letting both λ_i and λ_f vary to solve for the Scattering cross section for all (ω, K) .
2. Keeping λ_i fixed (using a projection as in (4.23)) but letting both λ_f and θ_i vary to solve for the Scattering cross section for all (ω, K) .
3. Keeping both θ_i and λ_i fixed (again using a projection) to solve for the Scattering cross section for a one-dimensional subset of (ω, K) .

A rough outline of how these methods work is given in Figure 4.13. At the time of writing only the last approach has been used for the Cambridge ${}^3\text{He}$ spin-echo apparatus as it has been the most convenient for quickly resolving the SF and ISF. However, this technique has significant theoretical issues that impair a true reconstruction of the ISF, hence the mention of two-dimensional alternatives.

Method 1: We fix θ_i for now and write $\rho(\lambda_i, \lambda_f)$ like before. In this case we transform from (λ_i, λ_f) to (K, ω) using the formulae (4.49), where we call this transformation $\Phi(\lambda_i, \lambda_f) = (K(\lambda_i, \lambda_f), \omega(\lambda_i, \lambda_f))$. Recall that the original definition (4.44) of the scattering cross section is per unit flux of incoming scattering particles. Therefore,

one must first correct for the spread of incoming wavelengths like so

$$\tilde{\rho}(\lambda_i, \lambda_f) = \frac{\rho(\lambda_i, \lambda_f)}{\int_{\mathbb{R}} \rho(\lambda_i, \lambda_f) d\lambda_f}. \quad (4.50)$$

Of course, one must be aware that the denominator here will lead to blow up of existing noise if the flux of the ${}^3\text{He}$ beam is very small for a given initial wavelength. Therefore, one should only work with a range of λ_i where the total intensity $\int_{\mathbb{R}} \rho(\lambda_i, \lambda_f) d\lambda_f$ is significantly larger than the noise level. For this reason it may be preferable to work with a *wider* range of initial wavelengths rather than a highly focused one.

Next, since $\rho(\lambda_i, \lambda_f)$ denotes a *Wavelength Intensity* function, if we wish to convert this to a *Momentum/Energy Intensity* function, say $\tilde{S}(K, \omega)$, then one must compensate for the non-linear transform given in (4.49) through a Jacobian term in order to preserve the notion of intensity:

$$\tilde{S}(K(\lambda_i, \lambda_f), \omega(\lambda_i, \lambda_f)) = \tilde{\rho}(\lambda_i, \lambda_f) \cdot (|\det(D\Phi)|(\lambda_i, \lambda_f))^{-1}. \quad (4.51)$$

From here $\tilde{S}(K, \omega)$ can be identified as being proportional to the scattering cross section and \tilde{S} can be viewed proportional to S in (4.46).

Method 2: This time we fix λ_i using a Fourier slice projection so now we write $\rho(\lambda_f, \theta_i)$. Since there is now only a fixed value of λ_i , no normalisation of the form (4.50) has to be carried out. However one still has to apply the same Jacobian technique as in (4.51):

$$\tilde{S}(K(\lambda_f, \theta_i), \omega(\lambda_f, \theta_i)) = \rho(\lambda_f, \theta_i) \cdot (|\det(D\Phi)|(\lambda_f, \theta_i))^{-1}, \quad (4.52)$$

where $\Phi(\lambda_f, \theta_i) = (K(\lambda_f, \theta_i), \omega(\lambda_f, \theta_i))$ where K and ω are defined by (4.49). While this approach is cleaner than method 1 since one does not need (4.50), one disadvantage is that one is forced to interpolate between different values of θ_i while for method 1 this can be avoided by taking advantage of the two-dimensional Fourier structure that defines $\rho(\lambda_i, \lambda_f)$ (see Section 4.4.4).

Method 3: Here $\rho(\lambda_f)$ depends only on the outgoing wavelength with λ_i, θ_i fixed. This is essentially the same as method 2 but now we only know $\tilde{S}(K, \omega)$ on a one-dimensional path in (K, ω) space. Because of this issue, some choose to exclusively work with the energy ω . With this approach (4.52) must be modified to work with intensity in ω only:

$$\tilde{S}(\omega(\lambda_f)) = \rho(\lambda_f) \cdot \left(\frac{d\omega}{d\lambda_f}(\lambda_f) \right)^{-1}. \quad (4.53)$$

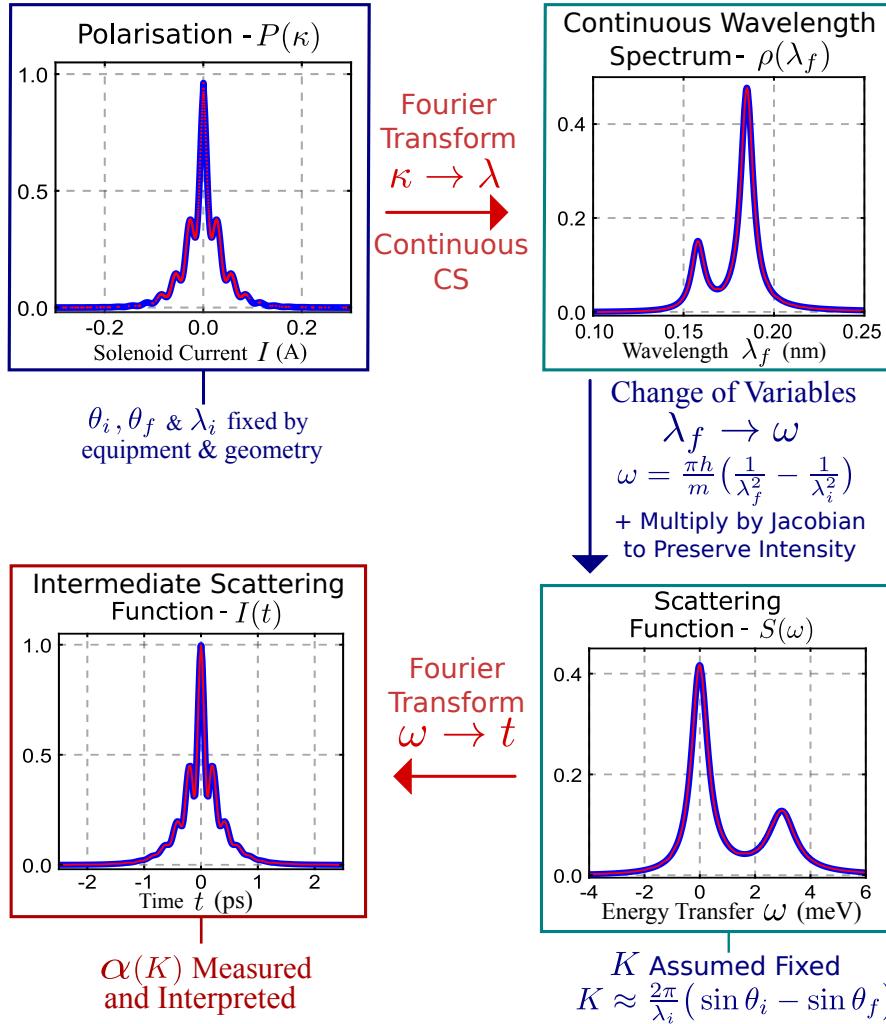


Figure 4.14: A detailed 1D description of the full loop under method 3 (One dimensional approach), along with continuous CS reconstructions. Blue plots denote the true underlying signal and red plots denote the reconstructions (except for the top left where the red dots denote the sampling points).

From here one may proceed, as in [58], to Fourier transform \tilde{S} to derive an approximation to the ISF I in (4.45). One problem however, is that the paths taken in (K, ω) space with θ_i, λ_f fixed are not straight lines parallel to the K -axis which otherwise would have justified this method using the Fourier slice theorem. Because of this issue we refer to $I(t)$ here as an *approximate-ISF*. Regardless, this method is arguably the quickest of all three, since a minimal number of one-dimensional polarisation scans is required to get this approach up and running. A picture demonstrating method 3 for a model problem is given in Figure 4.14.

4.4.4 Why One Should Use Continuous CS for Diffusion

We have just discussed how to go from polarisation data to the ISF. Next we discuss how to compress this process.

As Figure 4.13 shows, there are many steps that have to be taken in order to take polarisation data and use it to resolve the ISF/correlation functions. This is a considerably more complex problem than simply inverting a two dimensional Fourier transform like in the phonon problem. Now one has to consider what happens to the data after the initial Fourier transform step.

For example, suppose that one has reconstructed the wavelength intensity function $\rho(\lambda_i, \lambda_f)$ using a discrete Fourier transform. This means that we recover the truncated Fourier series approximation of ρ evaluated at a discrete grid of points:

$$\boldsymbol{\lambda}_{j,N} = (\lambda_{j_1,N}, \lambda_{j_2,N}), \quad \lambda_{j,N} = a + \frac{b-a}{2N+1}j, \quad j = 0, \dots, 2N. \quad (4.54)$$

If one then proceeds with a coordinate transformation (4.49) to solve for the SF $S(K, \omega)$ then the uniform grid of points in (λ_i, λ_f) becomes a non-uniform grid of (K, ω) . If one then wants to proceed with reconstructing the ISF or correlation function, then one must somehow invert a Fourier transform from knowledge of Fourier data at non-uniform points. There are many ways of tackling this issue, for example using some sort of interpolation to extend to a uniform grid.

However a very simple solution is to note that the original Fourier series approximation (1.11) of the wavelength intensity function $\tilde{\rho}_N(\lambda_i, \lambda_f)$ can be evaluated at any pair of points (λ_i, λ_f) in the region of interest. This means that we do not have to worry about handling a non-uniform grid. Even so, one issue with using a continuous Fourier series approximation lies with the topic of compression. Recall that the original compressed sensing approach for phonons relied on using a *discrete* Fourier transform matrix to solve for a *discrete* set of points, which is incompatible with using a *continuous* Fourier series approximation.

This is precisely where continuous CS comes in handy. The continuous CS approach in (4.32) aims to reconstruct a continuous version of the wavelength intensity function without the need of discrete grids. By providing an intrinsically continuous solution it fits into this multi-step framework presented in Figure 4.13 without adding further complication due to gridding. The scattering function can be evaluated at a particular point in momentum-energy space K, ω by finding the corresponding $(\lambda_i, \lambda_f, \theta_i)$ to evaluate the wavelength intensity function for and use the appropriate Jacobian transformation (e.g.

(4.52). Figure 4.14 is a proof of concept of this approach.

Consequently, we see that continuous CS is the *natural* choice for the diffusion problem as it simplifies the whole process both experimentally and theoretically; problems surrounding gridding/interpolation are absent⁵ and, in the case of methods 1 and 2, inaccuracies due to Fourier slicing are absent as there is no longer any projections.

⁵Except for, in the case of method 2, the possible interpolation in θ .

Chapter 5

Conclusions & Future Work

5.1 Developments with Compressed Sensing

5.1.1 Tools for Identifying Good Reconstruction Bases

A large part of this thesis focused on the study of line/block coherences in various inverse problems of Fourier type. In the one-dimensional case, it was shown that Daubechies wavelets give optimal asymptotic incoherence decay rates of $\Theta(N^{-1})$ while for the polynomial case the optimal Fourier decay rates were $\Theta(N^{-2/3})$. Not only did this suggest that using polynomials for CS was relatively ineffective, work on optimality in Section 2.5 showed that the wavelet case attained the theoretical limit for all reconstruction bases as a power law. Therefore, we can derive the following heuristic:

- 1). When working with a Fourier sampling basis and a reconstruction basis to form a change of basis matrix U , decay of $\mu(R_N U), \mu(UR_N) = \Theta(N^{-1})$ should exist to allow for good subsampling *under the presence of asymptotic sparsity*. When this does not hold, *poor subsampling results should be expected* unless the benefits of asymptotic sparsity using the reconstruction basis completely overwhelms this lack of decay.

However, this heuristic does not give the complete picture. Not only should a reconstruction basis have optimal decay of $\Theta(N^{-1})$, it should also have a *practical optimal ordering of the Fourier basis*. In the one dimensional case this was always satisfied since every optimal ordering of the Fourier basis was by size of frequency. However, in two dimensions we saw that there were many differing types of optimal orderings (Linear, Hyperbolic, Semihyperbolic).

In practice one typically uses sampling schemes based on a linear scaling scheme [3.3.4](#) which means that the optimal orderings should be consistent with these approaches. Furthermore, there are physical reasons behind why such shapes are used. For example in HeSE the currents (I_i, I_f) generated by the two power supplies corresponds to the frequencies that are sampled. For a hyperbolic ordering, one would be required to use very large currents in one of the coils instead of spreading the balance across both coils. For example suppose the power supplies were set to work in current steps of $0.01A$ each. Then this means sampling a current of $100A$ in the incoming coil and $0A$ in the outgoing coil would have the same sampling priority as $1A$ across both coils simultaneously. Compare this to the linear case where $1A$ across both coils would have a similar sampling priority to sampling $1A$ from just one coil.

This leads us to the second heuristic (with a caveat):

- 2).** The optimal ordering of the Fourier basis should be linear unless both the underlying signal structure is made up of tensor products and the signal measurement process can sample relatively large frequency ranges in each dimension separately with little cost.

Consequently, these heuristics can be used to not only identify which bases are potentially good for CS based reconstructions (e.g. separable wavelet bases with sufficient smoothness), but they also rule out others. For example, the 3D separable Haar wavelet basis has inferior decay $\mu(R_N U) = \Theta(N^{-2/3})$ with a linear ordering of the Fourier basis (see Example [3.3.13](#)).

In fact Section [3.3.7](#) showed that a linear ordering cannot be optimal with decay $\mu(R_N U) = \Theta(N^{-1})$ without there being sufficient smoothing. With this in mind, it is no surprise that higher dimensional reconstruction elements such as shearlets (with established optimal approximation results for 2/3D signal classes [\[69\]](#)) demand relatively strong Fourier transform decay conditions in their construction.

5.1.2 Directions Forward for Local Coherence Analysis

The are a few potential directions that infinite dimensional local coherence theory could be taken in. Naturally, other reconstruction bases could be studied. In fact other reconstruction schemes such as frames can be analysed using a generalisation of the current approach. Instead of working with maxima over rows/columns of a change of basis matrix one would work with suprema of Fourier transforms (and the analogue of

the wavelet transform for the chosen reconstruction framework). It has been made aware to the author that such work has already begun for Shearlets with numerical evidence suggesting linear ordering-based decay for the 2D case.

One specific theoretical question left open is the fastest decay of $\mu(R_N U)$ in general where both U and $U^\dagger = U^{-1}$ are both isometries. Recall that we have derived tight results on the decay for the general case where U is just an isometry (Theorems 2.6.1 and 2.6.3) however for many cases the two bases span the same space which means the change of basis matrix is invertible. Whether $\mu(R_N U) = \Theta(N^{-1})$ is the best possible decay (and not just as a power of N) in this case is currently unknown.

Of course it is not claimed that studying optimal decay of line/block coherences is the only direction to take the study of local coherence. As we have seen when establishing the $s \log(N)$ sampling rate for 1D wavelets, we will often need other tools to give a full picture of coherence behaviour. Not to mention that an approach that integrates with asymptotic sparsity for particular use cases would be beneficial, although there has already been some work on this area [15].

5.2 Developments with Helium Atom Scattering

This is a good time to have a look at what has been done already and what is yet to be done with CS and HeSE.

Firstly we have the work on phonons that has had the most development on the experimental side. Applying CS to this setting was a relatively straightforward application of tools originally designed for NMR and MRI type problems. Although there was the additional complication of the Fourier slice technique, this had already been tackled by previous work in the SMF research group [70] and was already aware to the AHFA group thanks to its application to the Radon transform [71]. The next real challenge was to deal with signal types that differ drastically to that of typical MRI problems, as can be seen from Figures 4.9 and 4.10 where there are a mixture of smooth and Dirac delta-like features. While some considerable compression (around 20%) has been achieved in these settings, further work on how best to compress these types of signal could lead to further advances.

What has been less explored, at least experimentally, is the application of continuous CS for detecting molecular diffusion using the full loop as described in section 4.4.3. An example of the current state of this collaboration is shown in Figure 4.1. The figure illustrates a 2D synthetic model of what we would like achieve with real data. Observe

that the reconstructions here are actually in 1D, therefore we are still working with method 3 in section 4.4.3.

After discussions with the SMF group, the proposed future plan of action is as follows:

1. First implement method 3 on the new power supplies and compare with the approach of taking direct measurements using a linear approximation [18, p. 9] [70, p. 13].
2. Implement method 2. Experimentally this is essentially the same as before, except it involves changing the incoming scattering angle. The challenge here is primarily on the numerical side as the data has to be interpolated with respect to theta.
3. Identify the feasibility of implementing method 1. In particular identify whether broadening the incoming wavelength profile is feasible. Theoretically, this is the simplest of the methods as the solution is now a continuous two-dimensional function.

The pay-off regarding method 3 requires some elaboration. Although, by broadening the incoming wavelength spectra, we potentially lose some energy resolution, however, we will be able to cover more of energy-momentum space for the Scattering function whilst removing any inaccuracies due to convolution with the incoming wavelength distribution that is intrinsic to the Fourier slice methods 2 and 3.

It is also worth mentioning that all these methods offer approaches that resolve both phonon and diffusion data simultaneously. This unified treatment of both problems is another advantage over the direct method in [18].

Appendix A

Hyperbolic Orderings in the Separable Case

Proof of Proposition 3.3.6: (3.36) applied to part 1.) of Lemma 3.1.3 shows that the decay of $\mu(\pi_N U)$ is bounded above by¹ $F_{\text{hyp}}(\sigma(N)) = 1/H_d(\sigma(N)) \approx 1/h_d(N)$, which gives us the upper bound for $\mu(P_N U)$ since $1/h_d(N)$ is decreasing.

For the lower bound, we focus on terms of the form $\lambda_d \circ \rho(m) = (t, \dots, t)$ for some $t \in \mathbb{N}$ and we set, for a fixed $q \in \mathbb{N}$

$$s = (1, \dots, 1), \quad j := \lceil \epsilon \log_2 t \rceil + q,$$

where we assume for now that $j \geq J$ is satisfied. This gives us

$$\begin{aligned} |\langle \Psi_{j,0}^s, \rho(m) \rangle|^2 &= \epsilon^d 2^{-dj} \prod_{i=1}^d |\mathcal{F}\psi(\epsilon 2^{-j} t)|^2 \\ &\geq \frac{1}{2^{d(1+q)} t^d} \cdot |\mathcal{F}\psi(\epsilon 2^{-(\lceil \epsilon \log_2 t \rceil + q)} t)|^2 \\ &\geq \frac{1}{2^{d(1+q)} t^d} \cdot L_q^{2d} \quad (\text{using (2.2.4)}). \end{aligned} \tag{A.1}$$

Let m now be arbitrary with $\prod_{i=1}^d \max(|\lambda_d \circ \rho(m)_i|, 1) = M \geq 1$ and let $t = \lceil M^{1/d} \rceil + 1$. Because ρ corresponds to the hyperbolic cross there exists an $m' > m$ such that $\prod_{i=1}^d \max(|\lambda_d \circ \rho(m')_i|, 1) = t^d$ where $\lambda_d \circ \rho(m') = (t, \dots, t)$. Notice that $t^d \leq E(d)M$ for some constant dependent on the dimension d . Furthermore, (A.1) holds for $m = m'$ if we have that $j \geq J$, which is satisfied if m is sufficiently large. Therefore we deduce by

¹for the definitions of H_d, h_d see (3.20) and (3.8).

(A.1) that

$$\begin{aligned}
\mu(R_m U) &\geq |\langle \Psi_{j,0}^s, \rho(m') \rangle|^2 \geq \frac{1}{2^{d(1+q)} t^d} \cdot L_q^{2d} \\
&\geq \frac{1}{E 2^{d(1+q)+1} M} \cdot L_q^{2d} \\
&= \frac{1}{E 2^{d(1+q)} \prod_{i=1}^d \max(|\lambda_d \circ \rho(m)_i|, 1)} \cdot L_q^{2d} \\
&\geq \frac{C}{E 2^{d(1+q)} h_d(m)} \cdot L_q^{2d} \quad (\text{using (3.21), } C > 0 \text{ some constant}).
\end{aligned}$$

This proves the lower bound.

Bibliography

- [1] A. D. Jones, B. Adcock, and A. C. Hansen, “On asymptotic incoherence and its implications for compressed sensing of inverse problems,” *IEEE Transactions on Information Theory (Accepted)*, 2015.
- [2] A. D. Jones, B. Adcock, and A. C. Hansen, “Analyzing the structure of multidimensional compressed sensing problems through coherence,” (*Preprint*), 2015.
- [3] A. D. Jones, A. Tamtögl, I. Calvo-Almazán, and A. Hansen, “Continuous compressed sensing of inelastic and quasielastic helium atom scattering spectra,” *PNAS (Submitted)*, 2015.
- [4] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [5] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [6] E. J. Candès, “An introduction to compressive sensing,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [7] 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.
- [8] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.
- [9] M. Fornasier and H. Rauhut, “Compressive sensing,” in *Handbook of Mathematical Methods in Imaging*, pp. 187–228, Springer, 2011.

- [10] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.
- [11] B. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Computing*, vol. 25, no. 2, pp. 227–234, 1995.
- [12] E. van den Berg and M. P. Friedlander, “Probing the pareto frontier for basis pursuit solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.
- [13] E. van den Berg and M. P. Friedlander, “SPGL1: A solver for large-scale sparse reconstruction,” June 2007. <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [14] E. J. Candès and Y. Plan, “A probabilistic and RIPless theory of compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.
- [15] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: A new theory for compressed sensing,” (*Preprint*), 2015.
- [16] M. Levitt, *Spin Dynamics, Basics of Nuclear Magnetic Resonance*. John Wiley & Sons, 2008.
- [17] D. J. Holland, M. J. Bostock, L. F. Gladden, and D. Nietlispach, “Fast multi-dimensional NMR spectroscopy using compressed sensing,” *Angewandte Chemie International Edition*, vol. 50, no. 29, p. 65486551, 2011.
- [18] F. Mezei, *Neutron Spin Echo*. Lecture Notes in Physics. Springer, 1980.
- [19] G. Alexandrowicz and A. P. Jardine, “Helium spin-echo spectroscopy: studying surface dynamics with ultra-high-energy resolution,” *J. Phys.: Condens. Matter*, vol. 19, 2007.
- [20] M. L. Dove, *Structure and Dynamics: An Atomic View of Materials*. Oxford University Press, 2002.
- [21] S. Mallat, *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third ed., 2009.
- [22] E. Hernández and G. Weiss, *A first course on wavelets*. Studies in Advanced Mathematics, Boca Raton, FL: CRC Press, 1996. With a foreword by Yves Meyer.

- [23] A. Cohen, I. Daubechies, and J. Feauveau, “Biorthogonal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, p. 485560, 1992.
- [24] G. Kutyniok and D. Labate, *Shearlets, Multiscale Analysis for Multivariate Data*. Birkhäuser, 2012.
- [25] C. Poon, “On the role of total variation in compressed sensing,” (*Preprint*), 2015.
- [26] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.
- [27] B. Adcock and A. C. Hansen, “Generalized sampling and infinite-dimensional compressed sensing,” *Foundations of Computational Mathematics*, 2015.
- [28] B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke, “Generalized sampling: extension to frames and inverse and ill-posed problems,” *Inverse Problems*, vol. 29, no. 1, p. 015008, 2013.
- [29] 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.*, vol. 182, p. 187279, 2014.
- [30] M. Gataric and C. Poon, “A practical guide to the recovery of wavelet coefficients from fourier measurements,” (*Preprint*), 2015.
- [31] B. Adcock, A. Hansen, and B. Roman, “A note on compressed sensing of structured sparse wavelet coefficients from subsampled fourier measurements,” (*Preprint*), 2014.
- [32] B. Adcock, A. Hansen, and B. Roman, “The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing,” (*Preprint*), 2014.
- [33] H. Rauhut. and R. Ward, “Sparse legendre expansions via ℓ_1 -minimization,” *Journal of Approximation Theory*, vol. 164, pp. 517–533, 2012.
- [34] J. Hampton and A. Doostan, “Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies,” *Journal of Computational Physics*, vol. 280, p. 363386, 2015.

- [35] F. Krahmer and R. Ward, “Stable and robust sampling strategies for compressive imaging,” *IEEE Trans. Image Process.*, vol. 23 (2), pp. 612–622, 2014.
- [36] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, “Realistic analytical phantoms for parallel Magnetic Resonance Imaging,” *IEEE Trans. Med. Imaging*, vol. 31, no. 3, pp. 626–636, 2012.
- [37] B. Adcock, A. Bastounis, A. C. Hansen, and B. Roman., “On fundamentals of models and sampling in compressed sensing,” (*Preprint*), 2015.
- [38] A. Cohen, I. Daubechies, and P. Vial, “Wavelet bases on the interval and fast algorithms,” *Journal of Applied and Computational Harmonic Analysis*, 1993.
- [39] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Comm. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.
- [40] G. N. Watson, *A Treatise on the Theory of Bessel Functions*. Cambridge, England: Cambridge University Press, 1944.
- [41] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, Washington, D.C., 1964.
- [42] L. J. Landau, “Bessel functions: monotonicity and bounds,” *J. London Math. Soc.* (2), vol. 61, no. 1, pp. 197–215, 2000.
- [43] K. I. Babenko, “On the approximation of functions periodic functions of several variables by trigonometric polynomials,” *Dokl. Akad. Nauk SSSR*, vol. 132, pp. 247–250, 1960.
- [44] R. A. Devor, P. P. Petrushev, and V. N. Temlyakov, “Multidimensional approximations by trigonometric polynomials with harmonics of a hyperbolic cross,” *Mat. Zametki*, vol. 56, no. 3, pp. 36–63, 158, 1994.
- [45] T. Körner, *Fourier Analysis*. Cambridge University Press, 1988.
- [46] C. Fellner, W. Müller, J. Georgi, U. Taubenreuther, F. Fellner, and W. Kalender, “A high-resolution phantom for MRI,” *Magn. Reson. Imaging*, vol. 19, pp. 899–904, 2001.

- [47] W. L. Bragg, “The diffraction of short electromagnetic waves by a crystal,” *Proceedings of the Cambridge Philosophical Society*, vol. 17, p. 4357, 1913.
- [48] G. L. Squires, *Introduction to the Theory of Thermal Neutron Scattering*. Cambridge University Press, 1978.
- [49] A. P. Jardine, G. Alexandrowicz, H. Hedgeland, W. Allison, and J. Ellis, “Studying the microscopic nature of diffusion with helium-3 spin-echo,” *Phys. Chem. Chem. Phys.*, vol. 11, no. 18, pp. 3355–3374, 2009.
- [50] U. Harten, J. P. Tonnies, and C. Woll, “Helium time-of-flight spectroscopy of surface-phonon dispersion curves of the noble metals,” *Faraday Discuss. Gem. Soc.*, vol. 80, pp. 137–149, 1985.
- [51] L. van Hove, “Correlations in space and time and born approximation scattering in systems of interacting particles,” *Phys. Rev.*, vol. 95, no. 1, p. 249, 1954.
- [52] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed Sensing MRI,” *IEEE Signal Process. Mag.*, vol. 25, pp. 72–82, March 2008.
- [53] C. Boyer, P. Weiss, and J. Bigot, “An algorithm for variable density sampling with block-constrained acquisition,” *Siam Journal of Imaging Sciences*, vol. 7, pp. 1080–1107, 2014.
- [54] J. V. Barth, “Transport of adsorbates at metal surfaces: from thermal migration to hot precursors,” *Surfaces Science Reports*, vol. 40, pp. 75–149, 2000.
- [55] A. Barreiro, R. Rurali, Eduardo, R. Hernndez, J. Moser, T. Pichler, L. Forr, and A. Bachtold, “Subnanometer motion of cargoes driven by thermal gradients along carbon nanotubes,” *Science*, vol. 320, pp. 775–778, 2008.
- [56] E. J. M. Daniel Dundas and and T. N. Todorov, “Current-driven atomic water-wheels,” *Nature Nanotechnology*, vol. 4, pp. 99–102, 2009.
- [57] P. Hofmann, *Surface Physics: An Introduction*. eBook 978-87-996090-1-7, 2013.
- [58] A. P. Jardine, H. Hedgeland, G. Alexandrowicz, W. Allison, and J. Ellis, “Helium-3 spin-echo: Principles and application to dynamics at surfaces,” *Prog. Surf. Sci.*, vol. 84, pp. 323–379, 2009.

- [59] F. Hofmann and J. P. Toennies, “High-resolution helium atom time-of-flight spectroscopy of low-frequency vibrations of adsorbates,” *Chem. Rev.*, vol. 96, pp. 1307–1326, 1996.
- [60] E. Hulpke, *Helium Atom Scattering From Surfaces*. Springer, 1992.
- [61] A. I. M. Rae, *Quantum Mechanics*. Taylor & Francis, 2008.
- [62] C. Kittel, *Introduction to Solid State Physics*. John Wiley & Sons Inc, 2005.
- [63] A. Lock, B. J. Hinch, and J. P. Toennies, *Kinetics of Ordering and Growth at Surfaces*. Springer, 1990.
- [64] E. M. McIntosh, P. R. Kole, M. El-Batanouny, D. M. Chisnall, J. Ellis, and W. Allison, “Measurement of the phason dispersion of misfit dislocations on the Au(111) surface,” *Phys. Rev. Lett.*, vol. 110, 2013.
- [65] P. R. Kole, *Dynamics and morphology of metal and metal oxide surfaces*. PhD thesis, University of Cambridge, 2011.
- [66] R. Bell, “Optimal sampling strategies for spin-echo experiments,” Master’s thesis, University of Cambridge, 2011.
- [67] O. Christensen, *Frames and Bases An Introductory Course*. Springer, 2008.
- [68] M. Bée, *Quasielastic Neutron Scattering*. CRC Press, 1988.
- [69] G. Kutyniok, J. Lemvig, and W. Q. Lim, “Optimally sparse approximations of 3d functions by compactly supported shearlet frames,” *SIAM J. Math. Anal.*, vol. 44, pp. 2962–3017, 2012.
- [70] G. Alexandrowicz, *Helium Spin Echo Spectroscopy: Measuring the Dynamics of Atoms, Molecules and Surfaces*. PhD thesis, University of Cambridge, 2005.
- [71] E. T. Quinto, “An introduction to X-ray tomography and Radon transforms,” in *The Radon Transform, Inverse Problems, and Tomography*, vol. 63, pp. 1–23, American Mathematical Society, 2006.