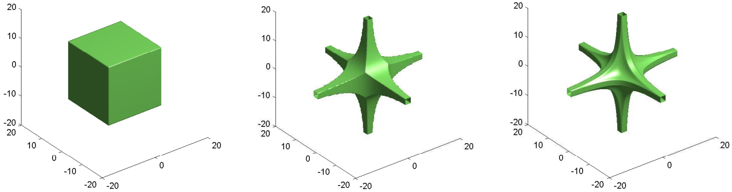
Matrix Suprema & Compressed Sensing

Alex Jones

Collaborative Work with Ben Adcock & Anders Hansen



Brief Outline of The Talk

1. Backgound: Inverse Problems and Compressed Sensing
2. General Theory: Incoherence, Orderings and Optimal Decay
3. One-Dimensional Examples & Tests
4. Multi-Dimensional Tensor Cases
5. Multi-Dimensional Separable Wavelet Cases
6. Multi-Dimensional Comparisons

Background: The Finite Dimensional Problem

Let us take two orthonormal bases $B_1 = (u_n)_{n=1}^N, B_2 = (v_n)_{n=1}^N$ of $\mathbb{C}^N$ and form the change of basis matrix $U \in \mathbb{C}^{N \times N}, U_{i,j} = \langle v_j, u_i \rangle$. B_1 is typically called the 'sampling basis' and B_2 the 'reconstruction basis'.

Suppose we have a vector $w \in \mathbb{C}^N$ that is 'simple' to express in basis B_2 but we can only recieve a small number $m \ll N$ of coefficients of the form $\tilde{w}_i := \langle w, u_i \rangle$ from B_1 (which we call *samples*). Is it possible to reconstruct w from such few coefficients?

The goal of compressed sensing, introduced by Candès, Donoho, Romberg, Tao et al., is to try and use the property that w is 'simple' when expressed in B_1 to somehow solve this seemingly ill-posed problem, with some caveats on w, m and the structure of U.

Background: Basic Concepts

What does 'simple' mean? We mean that the **sparsity** $s := \#\{v_i \in B_2 : \langle w, v_i \rangle \neq 0\}$ of non-zero coefficients is very small ($s \ll N$).

How do we take our samples? The m samples are taken **uniformly at random** from the set $(u_i)_{i=1}^{n}$.

What structure must U have? The matrix must have small **incoherence** $\mu(U) := \sup_{1 \leq i,j \leq N} |U_{i,j}|^2$ which can be interpreted as U being very spread out and flat. Ideally we would have $\mu(U) = 1/N$, in which case we say U is **perfectly incoherent**.

Background: Basic Concepts

How does this all fit together then? We expect (i.e. with high probability) a good reconstruction if

$$m \gtrsim \text{constant} \cdot \mu(U) \cdot s \cdot N \log(N)$$

How do we actually try to reconstruct w? We solve the convex optimisation problem (P_{samp} denotes the projection map onto the samples chosen)

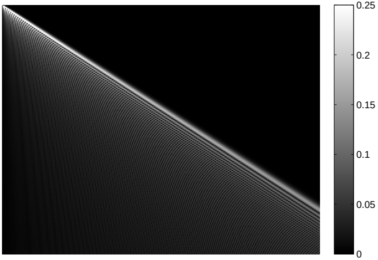
$$\hat{x} := \underset{x \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \|x\|_1 \quad \text{s.t.} \quad P_{\text{samp}} U x = P_{\text{samp}} \tilde{w} \right\}$$

This approach however does have its drawbacks as it assumes that $\mu(U)$ is small, which is typically an unreasonable assumption for large scale or infinite dimensional problems.

Background: Infinite-Dimensional Problems

Instead we typically have the behaviour $|U_{i,j}| \to 0$ as $i, j \to \infty$:

Figure: Fourier-Legendre Polynomial Matrix: Absolute Values



This suggests that we should not use a simple uniform approach, but instead rely on the structure of the problem. We also need to modify some of the concepts defined earlier.

Background: Infinite-Dimensional Bases

- **Fourier Basis** B_{f}: For $x \in \mathbb{R}$, define

$$\chi_k(x) \;=\; 2^{-1/2} \exp\left(2\pi \mathrm{i} k x\right) \;\cdot\; \mathbb{1}_{[-1,1]}(x), \qquad k \in \mathbb{Z}.$$

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

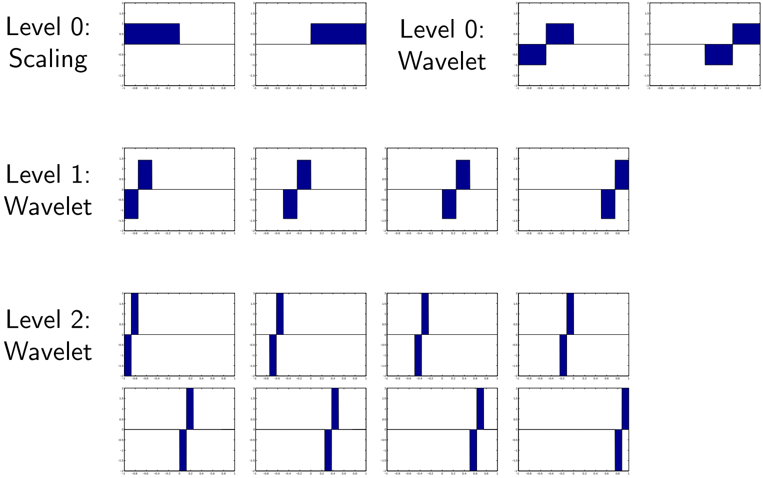
- **Legendre Polynomial Basis** B_{p}: $P_n(x)$ is an $(n-1)$-degree polynomial generated by the Gram-Schimdt othornormalisation procedure applied to the sequence $1, x, x^2, \ldots$ and the standard integral inner product on $L^2[-1,1]$.
- **Wavelet Basis** B_{w}: We use a scaling function ϕ and wavelet ψ and scale and shift them $(j \in \mathbb{N}, k \in \mathbb{Z})$

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

  We take all such functions whose supports overlaps with $[-1,1]$ to form the basis B_{w}.
  These two bases are often the reconstruction bases.

Background: Infinite-Dimensional Problems

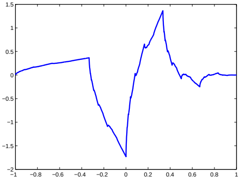
- Wavelets: These are a family of orthonormal functions that can be grouped into different 'resolution levels'. The most basic is probably the Haar wavelet basis which closely resembles pixel graphics:



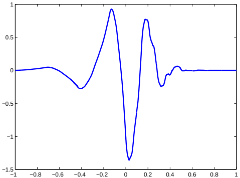
Background: Infinite-Dimensional Problems

There are many different types of wavelet apart from the Haar wavelet. One of the most famous types of wavelet are the Daubechies wavelets, which are indexed by by $n = 2, 4, 6, \ldots$. The higher n is the smoother the wavelet and there are other benefits involving sparse approximation, but the downside is they have larger support.

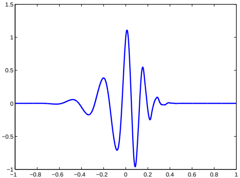
(a) Daubechies4



(b) Daubechies8



(c) Daubechies16



Concepts

In previous work by Adcock, Anders, Bogdan & Poon it was shown that the traditional compressed sensing concepts can be effectively generalised:

- Sparsity is changed to **sparsity in levels**. This means we break down $\mathbb{N}$ into regions S_i $(i = 1, ..., r)$ and define $s_i(f) := \#\{j \in S_i : \langle f, g_j \rangle\} \neq 0$ where $(g_j)_{j=1}^{\infty}$ is the basis B_2.
- **Subsampling** is also done **in levels**. Again, we break $\mathbb{N}$ (which denotes B_1 here) down into subsets Ω_i and uniformly subsample within these sets to different degrees.
- **Incoherence** is replaced by the **asymptotic incoherences** $\mu(P_N^{\perp} U), \mu(U P_N^{\perp})$ where $P_N^{\perp}(x) = (x_{N+1}, x_{N+2}, ...)$.

Subsampling Guarentees

Combining sparsity in levels $S_i = \{M_{i-1}+1, ..., M_i\}$ with subsampling in levels $\Omega_i = \{N_{i-1}+1, ..., N_i\}$ we require the number of samples m_i in Ω_i to satisfy the following if we expect good reconstruction:

$$m_k \gtrsim |\Omega_k| \cdot \sum_{l=1}^{r} \mu_{\mathbf{M},\mathbf{N}}(k,l) \cdot s_l \cdot \log(N_r),$$

where the local incoherence $\mu_{\mathbf{M},\mathbf{N}}(k,l)$ is defined by

$$\mu_{\mathbf{M},\mathbf{N}}(k,l) = \sqrt{\min\left((\mu(P^{\perp}_{N_k}U), \mu(UP^{\perp}_{M_l})\right) \cdot \mu(P^{\perp}_{N_k}U)}.$$

Therefore how much we can subsample in each level depends on how small the sparsity is and how small the asymptotic coherence is. (it should be mentioned that we have very much oversimplified things here as there are some others factors at play as well)

Start of the General Theory

The goal of this talk is to discuss the decay of $\mu(P_N^\perp U)$ as $N \to \infty$ in general and in specifc cases. We begin by reclarifying the theoretical framework:

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,

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

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

$U \in \mathcal{B}(\ell^2(\mathbb{N}))$ is then supposed to be 'the change of basis matrix from the basis B_2 to the basis B_1' and we are expected to study the decay of $\mu(P_N^\perp U)$ as $N \to \infty$.

A Slight Problem...Lost Without Orderings

From the way the problem is posed we expect this decay to depend only on the two bases but $\mu(P_N^{\perp} U)$ depends entirely on how we order the basis B_1. This forces us to make the following additional definitions:

Definition (Ordering)

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

Definition (Change of Basis Matrix)

For a basis pair (B_1, B_2), with corresponding orderings $\rho : \mathbb{N} \to B_1$ and $\tau : \mathbb{N} \to B_2$, form a matrix U by the equation

$$U_{m,n} := \langle \tau(n), \rho(m) \rangle. \tag{1}$$

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

Comparing Orderings and Decay Rates

We define the following linear projection operators from $\ell^2(\mathbb{N})$ to itself as follows:

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

$\mu(\pi_N U)$ is typically called the **row coherence** as it describes the maxima over the Nth row of U. We shall often be comparing it with the **asymptotic coherence** $\mu(Q_N U)$ (which is equal to $\mu(P_{N-1}^{\perp} U)$ for $N \geq 2$). For example we have the simple inequality $\mu(\pi_N U) \leq \mu(Q_N U)$.

Notice that if we permute the columns of U then this does not effect $\mu(Q_N U)$ or $\mu(\pi_N U)$, which means that $\mu(Q_N U)$ and $\mu(\pi_N U)$ are independent of the ordering of B_2.

Comparing Orderings and Decay Rates

At first glance there seems to be an extremely simple way to describe the fastest decay of $\mu(Q_N U)$:

Definition (Best ordering)

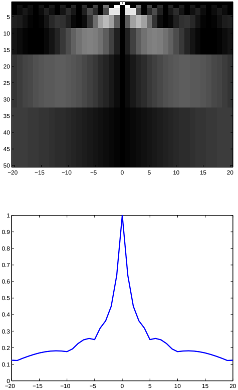
Let (B_1, B_2) be a basis pair. Then any ordering $\rho : \mathbb{N} \to B_1$ is said to be a 'best ordering' if for any other ordering τ of B_2 and $U = [(B_1, \rho), (B_2, \tau)]$ we have that the function $g(N) := \mu(\pi_N U)$ is decreasing.

While a best ordering certainly optimises the decay of $\mu(Q_N U)$, it turns out that this notion can lead to some unnecessarily complex orderings in many examples...

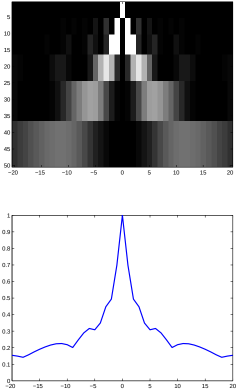
Comparing Orderings and Decay Rates

Here are two (20×20 centrally truncated) wavelet-Fourier Incoherence matrices 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).

Obeserve that the general decay behaviour is the same, even though the best orderings are not.



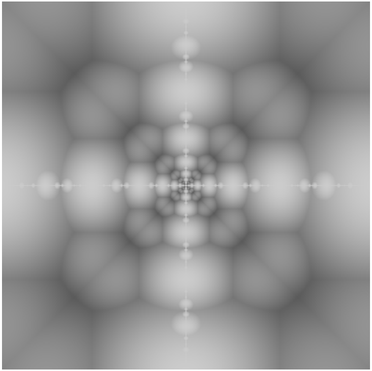
(d) Incoherence matrix and column maxima for a Haar wavelet basis.



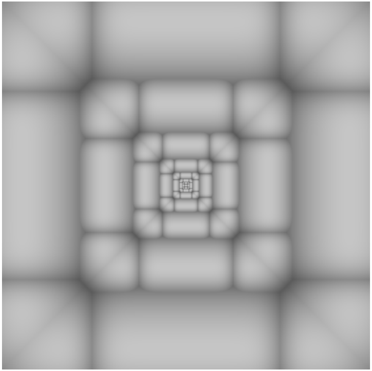
(e) Incoherence matrix and column maxima for Daubechies6 wavelet basis.

Comparing Orderings and Decay Rates

This seemingly minor difference only becomes more prominent in higher dimensions:



(f) 2D maxima for Haar wavelet basis.



(g) 2D maxima for Daubechies16 wavelet basis.

Comparing Orderings and Decay Rates

Definition (Relations on the set of orderings)

Let $\rho_1, \rho_2 : \mathbb{N} \to B_1$ be any two orderings of a basis B_1 and τ any ordering of a basis B_2. Let $U_1 := [(B_1, \rho_1), (B_2, \tau)]$, $U_2 := [(B_1, \rho_2), (B_2, \tau)]$. If there is a constant $C > 0$ such that

$$\mu(Q_N U_1) \le C \cdot \mu(Q_N U_2), \qquad \forall N \in \mathbb{N},$$

then we write $\boldsymbol{\rho_1 \prec \rho_2}$ 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 $\mathcal{R}(B_1)$.

Comparing Orderings and Decay Rates

Definition (Optimal ordering)

ρ is an **optimal ordering** for (B_1, B_2) if for every other ordering ρ' we have $\rho \prec \rho'$.

Definition (Optimal decay rate)

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

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

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

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

Comparing Orderings and Decay Rates

So how do we actually find optimal orderings and the optimal decay rate? The following tool often comes in handy:

Lemma

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

$$C_1 f(N) \leq \mu(\pi_N U) \leq C_2 f(N), \qquad \forall N \in \mathbb{N}, \tag{2}$$

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

If (2) holds for an ordering ρ then it is said to be a **strongly optimal ordering** for (B_1, B_2).

Theoretical Limits on the Decay

So how fast can the optimal decay get?

Theorem

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

In fact this result cannot be improved:

Lemma

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

$$\mu(Q_N U) \leq f(N), \qquad N \in \mathbb{N}. \tag{3}$$

If we restrict our decay function to be a power law, i.e. $f(N) := CN^{-\alpha}$ for some constants $\alpha, C > 0$ then the largest possible value of $\alpha > 0$ such that (3) holds for an isometry U is $\alpha = 1$.

Ordering the Bases

Apart from the Legendre polynomial basis, the other bases are currently unordered. We shall use the following tool to order bases in a straightforward fashion:

Definition (Consistent ordering)

Let $F : S \to \mathbb{R}$ where S is a set. We say that an ordering $\rho : \mathbb{N} \to S$ is 'consistent with respect to F' if

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

Definition (Standard ordering)

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

Ordering the Bases

Definition (Leveled ordering)

Define $F_{\mathrm{w}} : B_{\mathrm{w}} \to \mathbb{R}$ by

$$F_{\mathrm{w}}(f) = \begin{cases} j, & \text{if } f = \psi_{j,k} \\ -1, & \text{if } f = \phi_k \end{cases},$$

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

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.

Incoherence Results

Theorem

Let ρ be a standard ordering of B_{f}, τ a leveled ordering of B_{w} and $U = [(B_{\mathrm{f}}, \rho), (B_{\mathrm{w}}, \tau)]$. Then we have, for some constants $C_1, C_2 > 0$ the decay

$$C_1 \cdot N^{-1} \leq \mu(\pi_N U),\ \mu(U\pi_N) \leq C_2 \cdot N^{-1}, \qquad \forall N \in \mathbb{N}.$$

Consequently, both orderings are optimal and the optimal decays rates for (B_1, B_2) and (B_2, B_1) are both represented by the function $f(N) = N^{-1}$.

Incoherence Results

Theorem

Let ρ be a standard ordering of B_{f}, τ a natural ordering of B_{p} and $U = [(B_{\mathrm{f}}, \rho), (B_{\mathrm{p}}, \tau)]$. Then we have, for some constants $C_1, C_2 > 0$ the decay

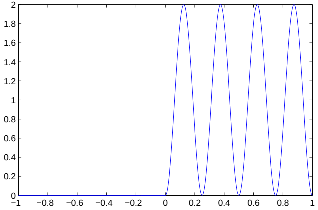
$$C_1 \cdot N^{-2/3} \leq \mu(\pi_N U),\ \mu(U \pi_N), \leq C_1 \cdot N^{-2/3}, \qquad \forall N \in \mathbb{N}.$$

Consequently, both orderings are optimal and the optimal decays rates for (B_1, B_2) and (B_2, B_1) are both represented by the function $f(N) = N^{-2/3}$.

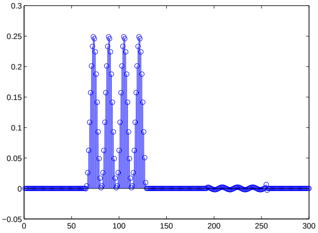
A Simple 1D Experiment

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

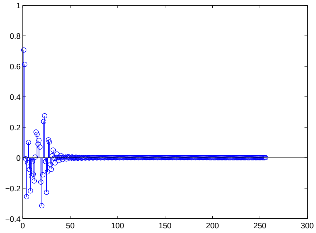
$$f(x) = (1 - \cos(8\pi x)) \cdot \mathbb{1}_{[0,1]}(x).$$



(a) Plot of f



(b) Wavelet Coefficients

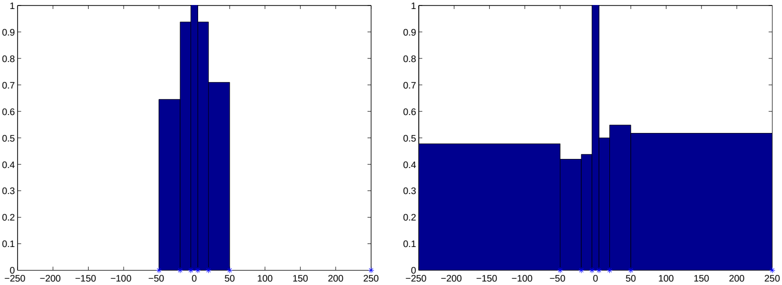


(c) Polynomial Coefficients

A Simple 1D Experiment

We shall be sampling Fourier coefficients and trying to reconstruct in Daubechies4 wavelets and in Legendre polynomials. We already know that there incoherence decays in a very different fashion (wavelets decay faster and therefore we should be able to subsample more). On the other hand polynomials provide a better direct approximation.

Figure: Two sampling patterns and their corresponding histograms.

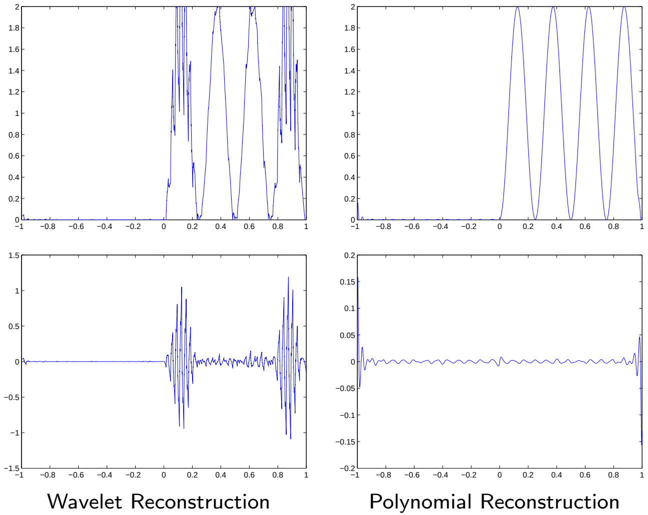


(a) Histogram for Pattern A

(b) Histogram for Pattern B

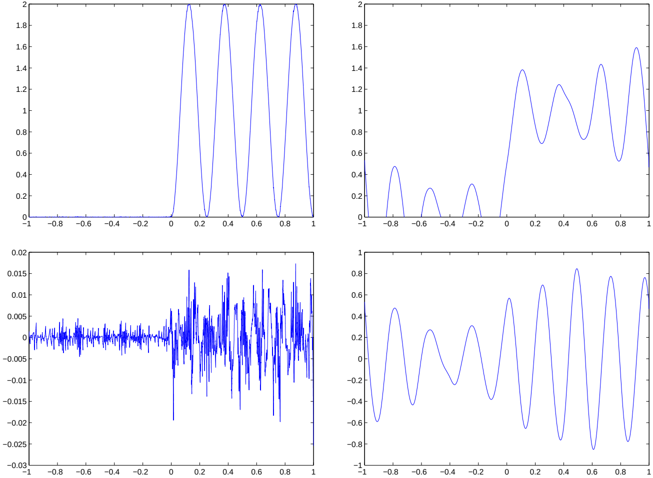
A Simple 1D Experiment

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



A Simple 1D Experiment

Figure: Reconstructions from Pattern B with errors.



Wavelet Reconstruction

Polynomial Reconstruction

Tensor Bases

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

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

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

$$\langle \rho^d(m), \rho^d(n)\rangle_{L^2(\mathbb{R}^d)} = \prod_{j=1}^d \langle \rho(m_j), \rho(n_j)\rangle_{L^2(\mathbb{R})}. \tag{4}$$

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 $d = 1$.

Tensor Bases

We would like to apply our results from the 1D case and extend them to cover the multidimensional tensor case:

Lemma

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

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

The Hyperbolic Approach

Suppose that we have a strongly optimal ordering ρ_1 of B_1 such that $f(n) = n^{-\alpha}$ for some $\alpha > 0$. The previous Lemma tells us that to find the optimal decay rate we should take an ordering $\sigma : \mathbb{N} \to \mathbb{N}^d$ 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 (Corresponding to the Hyperbolic Cross)

Let B_1^d be as before with corresponding d-dimensional indexing ρ_1^d induced by ρ_1. Define $F_H : \mathbb{N}^d \to \mathbb{R}$ by $F_H(n) = \prod_{i=1}^{d} n_i$. Then we say an ordering $\sigma : \mathbb{N} \to \mathbb{N}^d$ 'corresponds to the hyperbolic cross' if it is consistent with F_H.

The Hyperbolic Approach

Lemma (Hyperbolic Decay)

If $\sigma : \mathbb{N} \to \mathbb{N}^d$ *corresponds to the hyperbolic cross and* $d \geq 2$*, then*

$$\prod_{i=1}^{d} \sigma(N)_i \sim \frac{(d-1)!N}{\log^{d-1}(N+1)} =: h_d(N) \quad as \quad N \to \infty.$$

Definition (Hyperbolic Ordering)

If ρ_1 is a strongly optimal ordering for (B_1, B_2) then $\rho : \mathbb{N} \to B_1$ is said to be 'hyperbolic with respect to ρ_1' if we have

$$C_1 \cdot h_d(N) \leq \prod_{i=1}^{d} \left((\rho_1^d)^{-1} \circ \rho(N) \right)_i \leq C_2 \cdot h_d(N), \quad N \in \mathbb{N}.$$

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

The Hyperbolic Approach

This allows us to determine the optimal decay rate for when the optimal 1D decay rate is a power of N. First the Fourier-Wavelet case:

Theorem

Suppose that $B_1 = B_{\mathrm{f}}$, $B_2 = B_{\mathrm{p}}$, ρ_1 is a standard ordering and τ_1 is a natural ordering. 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, for some constants $C_1, C_2 > 0$,

$$\frac{C_1 \log^{d-1}(N+1)}{N} \leq \mu(\pi_N U_d),\ \mu(U_d \pi_N) \leq \frac{C_2 \log^{d-1}(N+1)}{N}, \qquad N \in \mathbb{N}.$$

If we compare this to our 1D result earlier we find that we gain extra log factors as we increase the dimension. Therefore, as the dimension increases, the optimal incoherence decay is getting worse and worse.

The Hyperbolic Approach

...and the Fourier-Polynomial case:

Theorem

Suppose that $B_1 = B_{\mathrm{f}}$, $B_2 = B_{\mathrm{p}}$, ρ_1 is a standard ordering and τ_1 is a natural ordering. 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, for some constants $C_1, C_2 > 0$, that for all $N \in \mathbb{N}$,

$$\frac{C_1(\log^{d-1}(N+1))^{2/3}}{N^{2/3}} \leq \mu(\pi_N U_d),\ \mu(U_d \pi_N) \leq \frac{C_2(\log^{d-1}(N+1))^{2/3}}{N^{2/3}}.$$

A Simple Hyperbolic Ordering

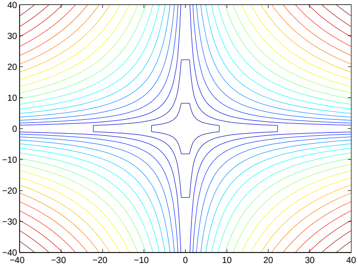
Example (Hyperbolic Cross in $\mathbb{Z}^d$)

Suppose that we define a function $F : \mathbb{Z}^d \to \mathbb{R}$ by

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

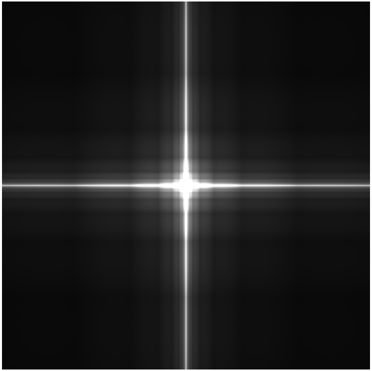
and say that a bijective function $\sigma : \mathbb{N} \to \mathbb{Z}^d$ 'corresponds to the hyperbolic cross in $\mathbb{Z}^d$' if it is consistent with F.

Figure: Hyperbolic Fourier Ordering in Two Dimensions

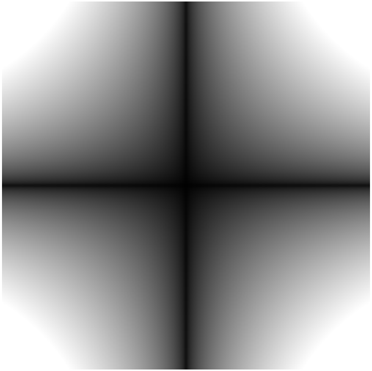


Tensor 2D Coherence

Figure: 2D Fourier-Haar Case



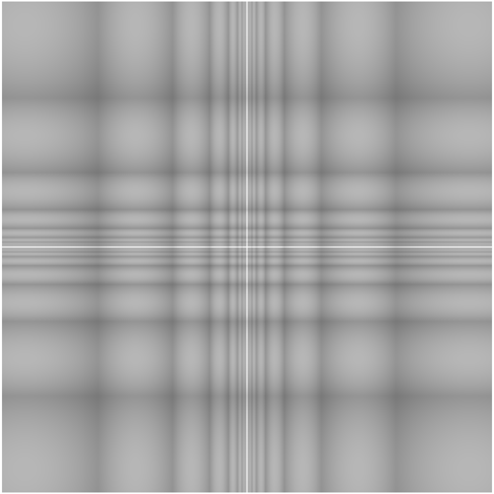
Original Coherence



Hyperbolic Scaling

Tensor 2D Coherence

Figure: 2D Fourier-Haar Case

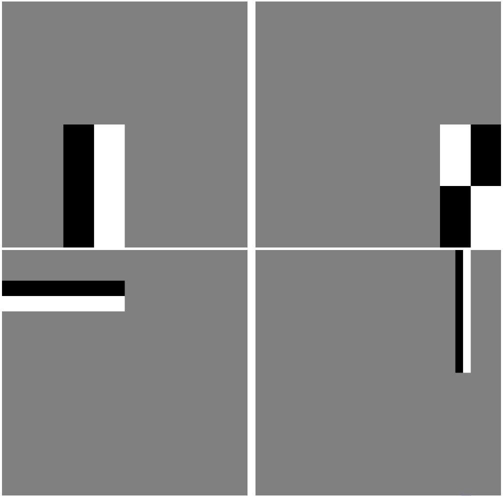


Scaled Coherence

Separable Wavelets - An Alternative to Tensors

While our argument works for all problems that involve a pair of tensor bases, this does not include the most widely used multidimensional wavelet basis - separable wavelets. Tensor wavelet bases have one major drawback: lack of a proper scaling levels

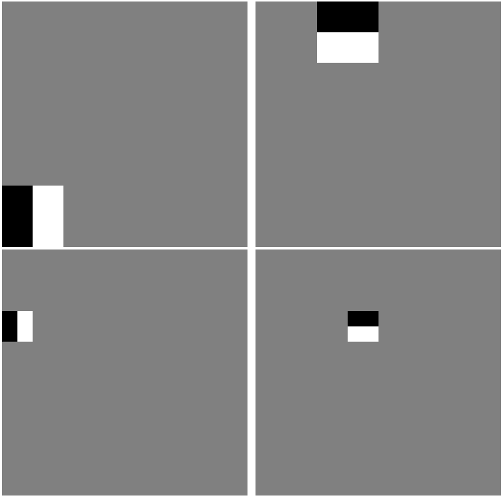
Figure: Elements from a 2D Haar Tensor Basis



Separable Wavelets - An Alternative to Tensors

Seprable wavelets, on the other hand, still have a resolution structure and as a result often provide better approximations to images.

Figure: Elements from a 2D Haar Separable Basis



Separable Wavelets - Definition

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

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

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

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

$\Phi(x)$ is used to define the resolution structure. Now 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 and define the functions

$$\Psi^s_{j,k} := \bigotimes_{i=1}^{d} \phi^{s_i}_{j,k_i}.$$

We then take all such functions whose support overlaps with $(-1,1)^d$ to form the basis B^d_{sep} (technically we also throw out functions with $s = 0$ and $j > J$).

Separable Wavelets - Leveled Ordering

Since we have a resolution structure we also have resolution levels and so...

Definition (Leveled Ordering)

For any $f \in B^d_{\text{sep}}$ define

$$F(f) = j \quad \text{if} \quad f = \Psi^s_{j,k}$$

Then we say that an ordering $\tau : \mathbb{N} \to B^d_{\text{sep}}$ is 'leveled' if τ is consistent with F.

Theorem

Let τ be any leveled ordering of B^d_{sep} and $U = [(B^d_{sep}, \tau), (B^d_{\mathrm{f}}, \rho)]$ for any ordering ρ of B^d_{f}. Then there are constants $C_1, C_2 > 0$ such that for all $N \in \mathbb{N}$ we have

$$\frac{C_1}{N} \leq \mu(\pi_N U) \leq \frac{C_2}{N}.$$

Therefore τ is strongly optimal for the basis pair $(B^d_{sep}, B^d_{\mathrm{f}})$.

...but how to order the Fourier Basis?

We form the d-dimensional (tensor) Fourier basis B_{f}^d by taking products:

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

It is also convenient to identify B_{f}^d with $\mathbb{Z}^d$ using the function

$$\lambda_d : B_{\mathrm{f}}^d \to \mathbb{Z}^d, \quad \lambda_d(\chi_k) := (\lambda(\chi_{k_1}), \dots, \lambda(\chi_{k_d})) = (k_1, \dots, k_d) = k.$$

This means we can view orderings on B_{f}^d as orderings on $\mathbb{Z}^d$. We already know an ordering on $\mathbb{Z}^d$, which is an ordering corresponding to the hyperbolic cross. So how does this turn out?

Trying a Hyperbolic Ordering

Proposition

Let $\sigma : \mathbb{N} \to \mathbb{Z}^d$ correspond to the hyperbolic cross in $\mathbb{Z}^d$ and define an ordering ρ of B_{f}^d by $\rho := \lambda_d^{-1} \circ \sigma$. Next let $U = [(B_{\mathrm{f}}^d, \rho), (B_{sep}^d, \tau)]$ for any ordering τ. Then there are constants $C_1, C_2 > 0$ such that for all $N \in \mathbb{N}$

$$\frac{C_1 \log^{d-1}(N+1)}{N} \leq \mu(Q_N U) \leq \frac{C_2 \log^{d-1}(N+1)}{N}.$$

Sadly we still have the extra log factors from the tensor case. Furthermore, since this estimate is for $\mu(Q_N U)$ and not $\mu(\pi_N U)$ it is not necessarily optimal, so we can possibly do better.

One natural option would be try and order $\mathbb{Z}^d$ according to some sort of measure of size. Maybe a norm will work?

Linear Orderings

Proposition

Let $\sigma : \mathbb{N} \to \mathbb{Z}^d$ correspond to a norm on $\mathbb{Z}^d$ and define an ordering ρ of B_{f}^d by $\rho := \lambda_d^{-1} \circ \sigma$. Next let $U = [(B_{\mathrm{f}}^d, \rho), (B_{sep}^d, \tau)]$ for any ordering τ. Furthermore, assume the following decay condition on the scaling function holds:

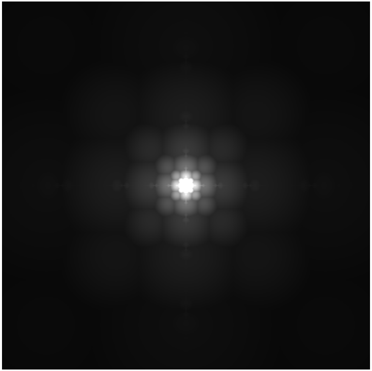
$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{d/2}}, \qquad \omega \in \mathbb{R} \setminus \{0\}, \tag{5}$$

where $\mathcal{F}$ denotes the Fourier Transform. Then there are constants $C_1, C_2 > 0$ such that for all $N \in \mathbb{N}$

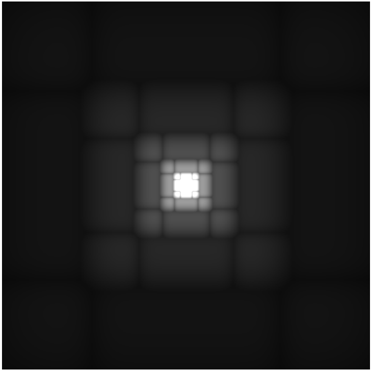
$$\frac{C_1}{N} \leq \mu(\pi_N U) \leq \frac{C_2}{N}.$$

This result tells that we can find strongly optimal orderings with the same decay rate as in 1D, provided that (5) holds, which is the case for all Daubechies wavelets in 2D.

2D Unscaled Incoherences

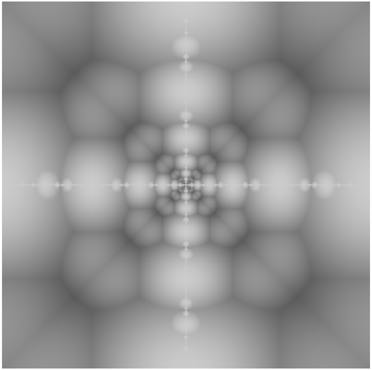


(a) 2D maxima for Haar wavelet basis.

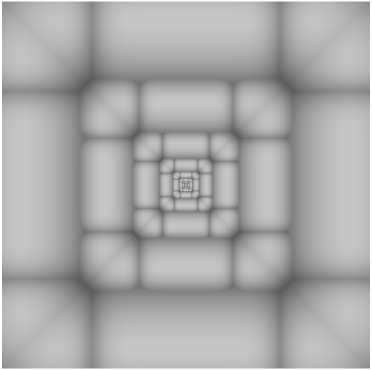


(b) 2D maxima for Daubechies16 wavelet basis.

2D Scaled Incoherences



(c) 2D maxima for Haar wavelet basis.



(d) 2D maxima for Daubechies16 wavelet basis.

Problems in Higher Dimensions

Example (3D Haar Wavelets)

If we do not have condition (5) then our argument can break down very badly: For Haar wavelets we have an explicit formula for the Fourier transform of the one-dimensional scaling function,

$$\mathcal{F}\phi(\omega) = \frac{\exp(2\pi i\omega) - 1}{2\pi i\omega}.$$

Therefore we have that (5) is not satisfied for $d = 3$. If ρ is chosen to be a linear ordering there are infinitely many m such that

$$|\langle \Phi, \rho(m) \rangle|^2 \geq \frac{E}{m^{2/3}},$$

for some constant E. Therefore an upper bound of the form Constant $\cdot N^{-1}$ is not possible for a linear ordering. Maybe we can try some kind of combination of a Linear and Hyperbolic ordering, but how?

Semi-Hyperbolic Orderings

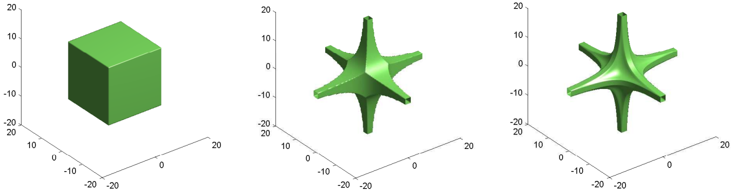
Definition

Let us define, for $r, d \in \mathbb{N}, r \leq d$ the function

$$H_{d,r}(n) := \max_{\substack{i_1,\ldots,i_r \in \{1,\ldots,d\} \\ i_1 < \ldots < i_r}} \prod_{j=1}^{r} \max(n_{i_j}, 1), \qquad n \in \mathbb{Z}^d.$$

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

Figure: Isosurfaces of $H_{3,r}$, $r = 1, 2, 3$ describing the three types of ordering in 3D



(a) $r = 1$ (Linear) ; Isosurface value=10.

(b) $r = 2$ (Semi-Hyperbolic) ; Isosurface value=20.

(c) $r = 3$ (Hyperbolic) ; Isosurface value=20.

Semi-Hyperbolic Orderings

Proposition

Let $\sigma : \mathbb{N} \to \mathbb{Z}^d$ *be semihyperbolic of order* r *in* d *dimensions (with* $r < d$*) and define an ordering* ρ *of* B_{f}^d *by* $\rho := \lambda_d^{-1} \circ \sigma$*. Next let* $U = [(B_{\mathrm{f}}^d, \rho), (B_{sep}^d, \tau)]$ *for any ordering* τ*. Furthermore, assume the following decay condition on the scaling function holds:*

$$|\mathcal{F}\phi(\omega)| \leq \frac{K}{|\omega|^{d/2r}}, \qquad \omega \in \mathbb{R} \setminus \{0\}, \tag{6}$$

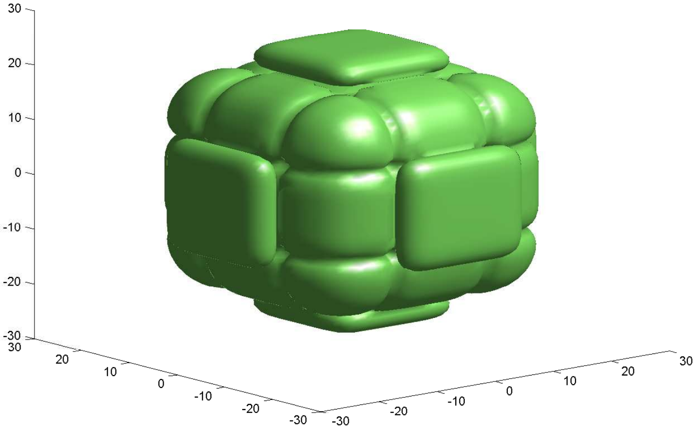
where $\mathcal{F}$ *denotes the Fourier Transform. Then there are constants* $C_1, C_2 > 0$ *such that for all* $N \in \mathbb{N}$

$$\frac{C_1}{N} \leq \mu(Q_N U) \leq \frac{C_2}{N}.$$

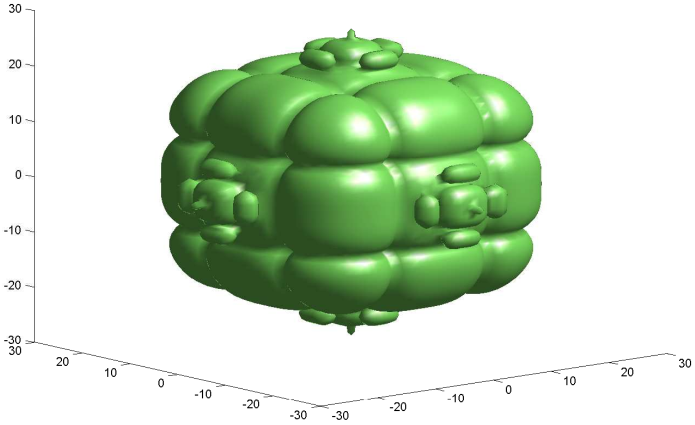
Furthermore, the ordering ρ *is optimal for the basis pair* $(B_{\mathrm{f}}^d, B_{sep}^d)$*.*

Given $d \in \mathbb{N}$ it is always possible to find an $r \in \{1, ..., d-1\}$ such (6) holds for any specific wavelet basis. Therefore we have found optimal orderings for any wavelet case in any dimension with decay as in 1D.

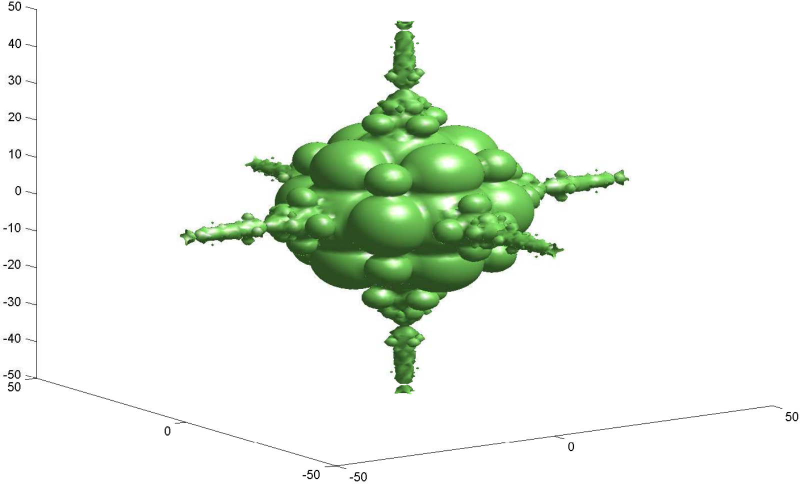
3D Incoherence Isosurfaces: Daubechies8



3D Incoherence Isosurfaces: Daubechies4



3D Incoherence Isosurfaces: Haar



A Final 2D Compressed Sensing Test

2D Haar Basis Incoherence Decay Rates		
Ordering	Tensor	Separable
Linear	$N^{-1/2}$	N^{-1}
Hyperbolic	$\log(N+1) \cdot N^{-1}$	$\log(N+1) \cdot N^{-1}$

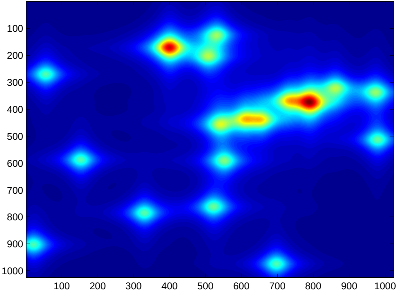
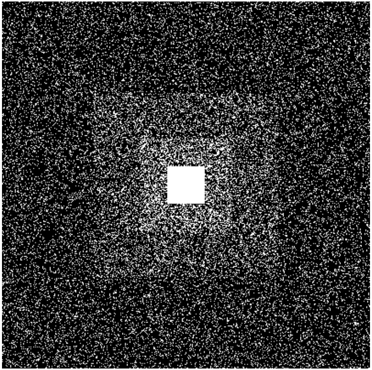
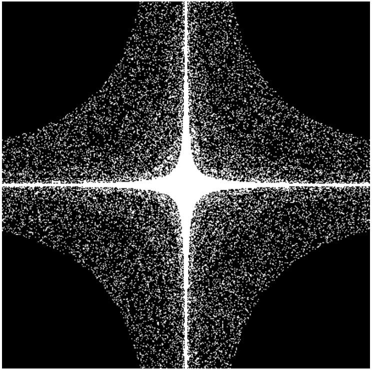


Figure: A 2D Spectrum that we'd like to reconstruct.

Sampling Patterns

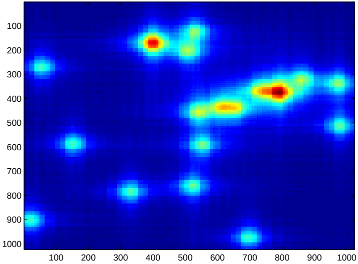


(a) Subsampling in Levels with a Linear Ordering

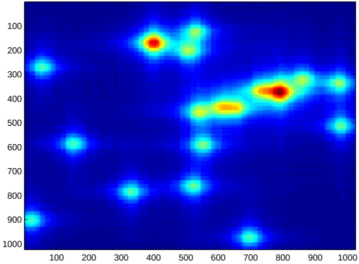


(b) Subsampling in Levels with a Hyperbolic Ordering

Reconstructions Using Pattern (a)

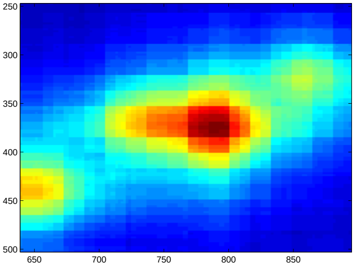


(c) Tensor Reconstruction

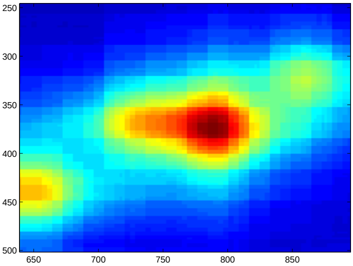


(d) Separable Reconstruction

Reconstructions Using Pattern (a)

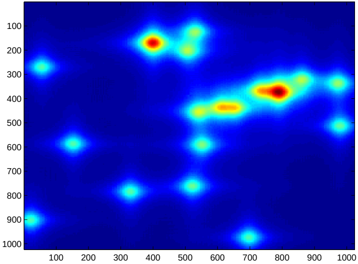


(e) Tensor Reconstruction (closeup)

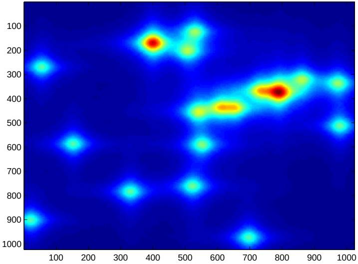


(f) Separable Reconstruction (closeup)

Reconstructions Using Pattern (b)

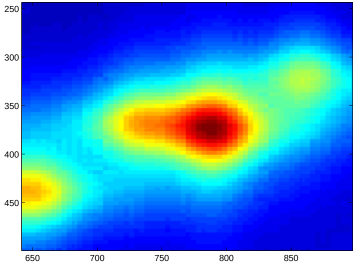


(g) Tensor Reconstruction

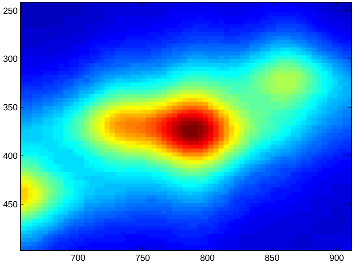


(h) Separable Reconstruction

Reconstructions Using Pattern (b)



(i) Tensor Reconstruction (closeup)



(j) Separable Reconstruction (closeup)

Possible Future Work

- A few questions still remain open
- 3D Tests!
- Tackling different examples e.g. frames
- Studying the local coherence in more detail
- Linking this in with problems with fixed sparsity strutures