Uncertainty Principles and Ideal Atomic Decomposition

David Donoho
Statistics Department
Stanford University

Xiaoming Huo
Industrial Engineering
Georgia Tech
Many Available Transforms

- Sinusoids
- Wavelets
- Cosine Packets
- Wavelet Packets

New Developments ...

- Anisotropic Wavelet Bases
- Ridgelets
- Curvelets
Lack of Universality

- There is no universal basis
- Different bases good for Different purposes

Example 1.

- Sinusoids for High-frequency oscillatory phenomena
- Wavelets for Impulsive phenomena
Example 2.

- Ridgelets for discontinuities along lines
- Wavelets for point discontinuities

Here good means **Sparsity**.
Combining Bases into Dictionaries

Mallat-Zhang (1993) Several Bases of interest

- Basis $\Phi_1$
- Basis $\Phi_2$
- ...$

Dictionary

$$\Phi = \{ \Phi_1 \cup \ldots \cup \Phi_d \}$$

Combined Representation

$$S' = \Phi \alpha \quad (= \sum_\gamma \alpha_\gamma \phi_\gamma)$$

Dream: If $S$ is the superposition of several ‘simple’ phenomena, $S_1, S_2, \ldots, S_d$, such that each $S_d$ is optimally represented in $\Phi_d$, we dream of representation of the superposed object which is the superposition of the optimal decompositions.
- \( \Phi \) overcomplete (several bases)
- Representation is non-unique
- Each \( S_d \) can be represented by any single one of the dictionaries.
- Representation of \( S_d \) by \( \Phi_{d'} \) is generally highly non-sparse.
Three Approaches

Method of Frames

\[(P_2): \quad \min ||\alpha||_2, \quad \text{s.t.} \quad S = \sum_\gamma \alpha_\gamma \varphi_\gamma,\]

\[||\alpha||_2^2 = \sum |\alpha_\gamma|^2.\]

Solvable by linear algebra \(\alpha = (\Phi'\Phi)^{-1}\Phi' S\).

Problem – generally completely dense.
Minimum $\ell^0$ decomposition

$$(P_0) : \quad \min \|\alpha\|_0, \quad \text{s.t.} \quad S = \sum_{\gamma} \alpha_\gamma \varphi_\gamma,$$

$\|\alpha\|_0 = \#\{\gamma : \alpha_\gamma \neq 0\}$ is the $\ell^0$ quasi-norm.

Search through $\Phi$ looking for a sparse subset providing exact decomposition.
Basis Pursuit

\[(P_1) : \min \| \alpha \|_1, \quad \text{s.t.} \quad S = \sum_{\gamma} \alpha_\gamma \varphi_\gamma,\]

\[\| \alpha \|_1 = \sum |\alpha_\gamma|.\]

Solvable by linear programming.

Modern interior-point optimization.

\(\ell^1\) is smallest convexification of \(\ell^0\) norm on set

\[\| \alpha_\infty \|_\infty \leq 1.\]
Experimental Results

Empirically, solution of BP is frequently quite sparse;

Many examples: Chen, Donoho, Saunders (1999)

Surprising Phenomenon

$S$ sparsely synthesized from only a few dictionary elements,

BP solution may *perfectly recover* the specific atoms and specific coefficients used in the synthesis.
Example

S. Chen (Thesis, 1996)

- $S$ a sum of 4 sinusoids and 2 spikes,

- Dictionary a combined time/frequency dictionary of sinusoids and spikes

- BP recovered exactly the indices and coefficients of the terms involved in the synthesis;

True across a wide range of amplitude ratios between the sinusoid and spike components.
Figure 1: Analyzing **TwinSine-1** with an overcomplete cosine dictionary.
Figure 2: Analysis versus synthesis of the signal **Carbon**.
Figure 3: Analyzing the signal **Carbon** with a wavelet packet dictionary.
Figure 4: MOF representation is not sparse.
Comparison

Matching Pursuit (Greedy) recovery of indices and coefficients was only approximate and became very inexact when the sinusoidal and spike components were at very different amplitudes.
Goal for Today

- In certain dictionaries $\Phi$
- Signal sufficiently sparse combination of terms from dictionary
- Unique Solution to $P_1$
- Unique Solution $P_0$
- Solutions are the same!

Refinements

- How much sparsity is required?
- What it is about dictionaries which makes this possible.
- What it is about $\ell^1$ which makes this possible.
Terminology

Any representation \( S = \sum_\gamma \alpha_\gamma \phi_\gamma \) is an atomic decomposition using atoms from the dictionary.

If

- \( S \) can be generated by highly sparse sum, (term “highly sparse” with appropriate definition), and
- there is only one highly sparse way to generate, and
- an optimization principle finds exactly that decomposition,

we say that the principle leads to ideal atomic decomposition under the stated sparsity hypothesis.
We claim that

- Under certain sparsity conditions,

- In certain dictionaries

- Minimum $\ell^1$-norm decomposition is an ideal atomic decomposition.
Combined dictionary: $\Phi = \Phi_1 \cup \Phi_2$

$\Phi_1$ spike basis

$$\varphi_{1,\tau}(t) = 1\{t=\tau\},$$

$\tau = 0, 1, \ldots, N - 1$

$\Phi_2$ Fourier basis

$$\varphi_{2,w}(t) = \frac{1}{\sqrt{N}} \exp(2\pi i wt/N)$$

$w = 0, 1, \ldots, N - 1.$

$\Phi_1, \Phi_2$ orthobases for $l^2_N$. 
Theorem 0.1  Let \( S = \sum_{\gamma \in T} \alpha_\gamma \varphi_\gamma + \sum_{\gamma \in W} \alpha_\gamma \varphi_\gamma \) where \( T \) is a subset of the “time domain” \( \{ (1, \tau) \} \) and \( W \) is a subset of the “frequency domain” \( \{ (2, w) \} \). If

\[
|T| + |W| < \sqrt{N},
\]

then \((P_0)\) has a unique solution. Meanwhile, there exist \((S, T, W)\) so that

\[
|T| + |W| = \sqrt{N}
\]

and \((P_0)\) has a non-unique solution.
**Theorem 0.2** Let $S = \sum_{\gamma \in T \cup W} a_\gamma \varphi_\gamma$ with $T, W$ as in Theorem ???. If

$$|T| + |W| < \frac{1}{2} \sqrt{N},$$

then $(P_1)$ has a unique solution, which is also the unique solution of $(P_0)$. Meanwhile, there exist $(S, T, W)$ so that

$$|T| + |W| = \sqrt{N}$$

and $(P_1)$ has a non-unique solution.

If the signal $S$ **truly** has a **very sparse** decomposition in the time/frequency dictionary, ..

- the solution is unique, and

- basis pursuit ($\ell^1$ decomposition) will find it.
Digression 1: Uncertainty Principle

Underlying Theorems is an uncertainty principle:

The *combined analysis* of a signal in the time and frequency domains cannot yield a transform pair which is sparse in both domains simultaneously.

Ideal atomic decomposition demands under sufficiently strict interpretation of the term ‘sparsity’,

*A signal cannot be sparsely synthesized from both the frequency side alone and from the time side alone at the same time.*

Otherwise, the atomic decomposition would be nonunique, so no ideal atomic decomposition.
New uncertainty principle, and generalization to other pairs than time-frequency.
Digression 2: Nonlinearity of $\ell^1$

Ideal atomic decomposition depends on very particular properties of the $\ell^1$ norm.

In effect, $(P_1)$ asks to find the member of a linear subspace closest to the origin in $\ell^1$ norm.

This closest point problem (which would be a linear problem in $\ell^2$ norm) is highly nonlinear in $\ell^1$ norm, and the nonlinearity is responsible for our phenomenon.
Precedent **Logan’s Phenomenon**.

- Decompose signal $S(t) = B(t) + N(t) - \Omega$-bandlimited function and impulsive noise;

- Find $\tilde{B}$ $\Omega$-bandlimited $L^1$ closest to $S$.

- Suppose product $|\Omega||\text{supp}(N)| < c$

- Perfect separation: $\tilde{B} = B$

- Phenomenon highly nonlinear in sense that perfect reconstruction holds at all signal/noise ratios.

Here bandlimiting $\leftrightarrow$ frequency domain sparsity.

Our generalization:

- Approximations in which time and frequency have symmetric role,

- No need for the frequency support of the signal to be an interval or even to be known.
Other Dictionary Pairs

**Theorem 0.3** Let $f(\theta)$ function on the circle $[0, 2\pi)$, a superposition of sinusoids and wavelets,

$$f(\theta) = \sum_{\lambda} \alpha_\lambda \psi_\lambda(\theta) + \sum_{n \geq n_0}^{\infty} c_n e^{in\theta}. \quad (1)$$

Here the $\psi_\lambda$ are the Meyer-Lemarié wavelets, and $n_0 = 2^{j_0+1}$. There is a constant $C$ with the following property. Let $N_j(\text{WAVELETS})$ be the number of Meyer Wavelets at resolution level $j$ and let $N_j(\text{SINUSOIDS})$ be the number of sinusoids at frequencies $2^j \leq n < 2^{j+1}$. Suppose that the sum obeys all the conditions

$$N_j(\text{WAVELETS}) + N_j(\text{SINUSOIDS}) \leq C \cdot 2^{j/2}, \quad (2)$$

$$j = j_0 + 1, \ldots$$
Consider the overcomplete dictionary $\Phi$ consisting of Meyer-Lemarié wavelets and of sinusoids at frequencies $n_0 \geq 2^{j_0+1}$. There is at most one way of decomposing a function $f$ in the form (??) while obeying (??). If $f$ has such a decomposition, it is the unique solution to the minimum $\ell^1$ optimization problem

$$\min \sum_{\lambda} |\alpha_\lambda| + \sum_{n \geq n_0}^\infty |c_n|.$$
Interpretation

No assumption about the sparsity or non-sparsity of the representation of $f$.

If sparsity holds, gives ideal atomic decomposition.

Notion of sparsity *level-dependent*.

Tolerate more nonzero terms at high resolution than we can at low resolution.

Little possibility of confusion between sparse sums of wavelets and sparse sums of sinusoids for sums limited to dyadic bands at high frequencies—two systems become highly disjoint.
Intuitively

A phenomenon near scale $2^{-j}$ and frequency $2^j$ cannot have sparse representation in wavelets basis and in sinusoid basis at the same time.

Formally

If a function $f$ has at most $C \cdot 2^{j/2}$ nonzero wavelet coefficients and sinusoid coefficients at each level $j$, then the function is zero.
Theorem [Hypotheses]

\[ f \in L^2(\mathbb{R}^2) \text{ a superposition of wavelets and ridgelets:} \]

\[ f = \sum_{Q} \alpha_Q \psi_Q + \sum_{\lambda \in \Lambda} \beta_{\lambda} \rho_{\lambda}. \]  \hfill (3)

\( \psi_Q \) are the usual two-dimensional Meyer-Lemarié wavelets for the plane.

\( \rho_{\lambda} \) are orthonormal ridgelets,

\( \Lambda \) consists of ridgelets at ridge scales \( j \geq j_0 + 2 \).

\( N_j(\text{WAVELETS}) \) number of wavelets at level \( j \),

\( N_j(\text{RIDGELETS}) \) number of ridgelets at level \( j \).
Theorem [Conclusion]

There is $C > 0$ so that if

$$N_j(\text{WAVELETS}) + N_j(\text{RIDGELETS}) \leq C \cdot 2^{j/2}, \quad (4)$$

$$j = j_0 + 2, \ldots$$

In the overcomplete dictionary $\Phi$ of Meyer-Lemarié wavelets and ridgelets with $\lambda \in \Lambda$, there is at most one way of decomposing a function $f$ in the form (??) which obeying (??).

Such a decomposition is the unique solution of the minimum $\ell^1$ optimization problem

$$\sum_Q |\alpha_Q| + \sum_\lambda |\beta_\lambda|.$$
Level-Dependent Sparsity

Tolerate more total terms at high resolution than at low resolution.

Little possibility of confusion between sparse sums of wavelets and sparse sums of ridgelets for sums from dyadic bands at increasingly high frequencies—the two systems become increasingly disjoint.
Intuitively,

A phenomenon occurring near scale $2^{-j}$ and frequency $2^j$ cannot have a sparse representation in both the wavelets basis and the ridgelets basis.

Formally,

If an $L^2$ function $f$ has at most $C \cdot 2^{j/2}$ nonzero wavelet coefficients and ridgelet coefficients at each level $j$, then the function is vanishing.
Agenda

- Time/Frequency: $\ell^0$
- Time/Frequency: $\ell^1$
- Appl: Bandlimited Approximation.
- From Complex to Real Sinusoids
- General Basis Pairs
- Appl: Speech Scrambling
- MultiScale Bases
\[ \ell^0 \text{ uncertainty principle} \]

**Theorem 0.4** Suppose \( (x_t)_{t=0}^{N-1} \) has \( N_t \) nonzero elements and that its Fourier transform \( (\hat{x}_w)_{w=0}^{N-1} \) has \( N_w \) nonzero elements. Then \( N_t N_w \geq N \) and so

\[ N_t + N_w \geq 2\sqrt{N}. \quad (5) \]

Donoho and Stark (1989)
Extremal functions \([N\) perfect square\]

\[
\text{III}_t = \begin{cases} 
1 & t = l\sqrt{N}, l = 0, 1, \ldots, \sqrt{N} - 1; \\
0 & \text{else}
\end{cases}
\]

and by its frequency and time shifts.

\text{III} is sparse

\[N_t + N_w = 2\sqrt{N}\]

\text{III} invariant under Fourier transformation:

\[\mathcal{F}(\text{III}) = \text{III}.\]
Negative Implication

III may equally be viewed as

(1) time domain synthesis using $\sqrt{N}$ spikes, or

(2) frequency-domain synthesis from $\sqrt{N}$ sinusoids.

$S = \text{III}, (P_0)$ has non-unique solution in overcomplete dictionary $\{\text{SPIKES}\} \cup \{\text{SINUSOIDS}\}$.

Conclusion: constraints on sparsity $N_t + N_w < K$ cannot guarantee uniqueness for $K > \sqrt{N}$. 
Positive Implication

Sparsity $N_t + N_w < K$, $K = \sqrt{N}$ can guarantee uniqueness.

Key Observation

Let $\mathcal{N} = \{ \delta : \Phi \delta = 0 \}$, For $\delta \in \mathcal{N}$, partition $\delta = (\delta^1, \delta^2)$ into components from two bases, $\delta \in \mathcal{N}$ implies

$$\Phi_1 \delta^1 + \Phi_2 \delta^2 = 0,$$

or

$$\delta^2 = -\Phi_2^T \delta^1.$$

$\mathcal{N}$ is the set of all pairs $\langle x, -\hat{x} \rangle$, where $x = (x_t)_{t=0}^{N-1}$ and $\hat{x} = (\hat{x}_w)_{w=0}^{N-1}$ is its Fourier transform.
Proof of Theorem 1

Suppose $S$ had two decompositions:

$$S = \Phi \alpha^1, \quad S = \Phi \alpha^2$$

$$0 = \Phi (\alpha^1 - \alpha^2).$$

so

$$\alpha^1 - \alpha^2 \in \mathcal{N}$$

Suppose both decompositions sparse:

$$\|\alpha^i\|_0 < \sqrt{N}.$$ 

Then there must be a transform pair $(x, -\hat{x})$ with $N_t + N_w < 2\sqrt{N}$. This would violate the $\ell^0$ uncertainty principle! Therefore $\ell^0$ optimization is unique if $S = \Phi \alpha$ and

$$\|\alpha\|_0 < \sqrt{N}.$$
2. Uniqueness of $\ell^1$ Optimization

For $\alpha$ unique solution, need

$$\|\tilde{\alpha}\|_1 > \|\alpha\|_1 \quad \Phi \tilde{\alpha} = \Phi \alpha$$

Equivalently, for every $\delta \in \mathcal{N}$ we must have

$$\|\alpha + \delta\|_1 - \|\alpha\|_1 > 0,$$

unless $\delta = 0$.

$$\|\alpha + \delta\|_1 - \|\alpha\|_1 = \sum_{(TUW)^c} |\delta_\gamma| + \sum_{TUW} (|\alpha_\gamma + \delta_\gamma| - |\alpha_\gamma|).$$

Now

$$\|\alpha + \delta\|_1 - \|\alpha\|_1 \geq \sum_{(TUW)^c} |\delta_\gamma| - \sum_{TUW} |\delta_\gamma|.$$
Sufficient condition for uniqueness: or \( \delta \neq 0 \),

\[
\sum_{T \cup W} |\delta_{\gamma}| < \sum_{(T \cup W)^c} |\delta_{\gamma}|, \quad \forall \delta \in \mathcal{N}. \quad (6)
\]

Less \( \ell^1 \) ON support than OFF support! Since \( \mathcal{N} = \) all pairs \((x, -\hat{x})\), condition is

\[
\sum_{T} |x_t| + \sum_{W} |\hat{x}_w| < \frac{1}{2} \left( ||x||_1 + ||\hat{x}||_1 \right), \quad (7)
\]

for every nonzero \( x \).
Time-frequency concentration measure

Given sets $T$ and $W$, define

$$
\mu(T, W) = \sup \frac{\sum_T |x_t| + \sum_W |\hat{x}_w|}{\|x\|_1 + \|\hat{x}\|_1},
$$

(8)

supremum over all nonzero $x = (x_t)_{t=0}^{N-1}$.

Uniqueness Condition

Uniqueness of $\ell^1$ optimization is implied by

$$
\mu(T, W) < \frac{1}{2}.
$$

(9)
Theorem 0.5  Let $T$ be a subset of the time domain and $W$ be a subset of the frequency domain. Then

$$\mu(T, W) \leq \frac{|T| + |W|}{\sqrt{N} + 1}.$$  \hspace{1cm} (10)

In particular, if $|T| + |W| \leq \frac{1}{2}\sqrt{N}$, then $\mu(T, W) < 1/2$, and the optimization problem $(P_1)$ has a unique solution.
**Proof**

**Lemma 0.6** Let \((x, \hat{x})\) be a Fourier transform pair. Then

\[ ||\hat{x}||_1 \geq \sqrt{N}||x||_{\infty}. \tag{11} \]

**Lemma 0.7** Consider the capacity defined by the optimization problem

\[
(K_{1,\tau}) \quad \min ||x||_1 + ||\hat{x}||_1, \quad \text{subject to} \quad x_\tau = 1.
\]

And the frequency-side capacity defined by the optimization problem

\[
(K_{2,w}) \quad \min ||x||_1 + ||\hat{x}||_1, \quad \text{subject to} \quad \hat{x}_w = 1,
\]

Then

\[
\text{Val}(K_{1,\tau}) = \text{Val}(K_{2,w}) = 1 + \sqrt{N}. \tag{12}
\]
Suppose

\[ S(t) = B(t) + \epsilon(t) \]

\( B \) discrete-time bandlimited signal with frequency-domain support purely in a certain \textit{unknown} band \( W \)

\( \epsilon \) is a discrete-time noise, of arbitrary size, supported in a set \( T \).

Suppose approximate \( S \) by minimal \( \ell^1 \) decomposition from combined time-frequency dictionary: (spikes, sinusoids).

Label \( \tilde{B} \) component of solution coming from sinusoids and

Label \( \tilde{\epsilon} \) component of \( \ell^1 \) solution coming from spikes, the approach

Bandlimited approximation with \textit{unknown} band \( W \)!
Our Results show

if $|\text{supp}(\hat{B})| + |\text{supp}(\epsilon)| \leq \frac{1}{2}\sqrt{N}$, then $\tilde{B} = B$ and $\tilde{\epsilon} = \epsilon$.

Other Sinusoids

\((\varphi_w)_{w=0}^{N-1}\) orthobasis for \(l_{2,N}\).

\(\tilde{x}_w = \langle x, \varphi_w \rangle\) Fourier-Bessel coefficients.

\(T\) and \(W\) subsets of \(t\)- and \(w\)- index space, respectively.

Define

\[
\tilde{\mu}(T, W; \varphi) = \sup \left\{ \frac{\sum_T |x_t| + \sum_W |\tilde{x}_w|}{\|x\|_1 + \|\tilde{x}\|_1} \right\},
\]

Earlier \(\mu(T, W)\) special case with \(\varphi_w = \frac{1}{\sqrt{N}} e^{i2\pi wt/N}\).
Put

$$\widetilde{M} = \max_w \max_t |\varphi_w(t)|,$$

Then

$$\widetilde{\mu}(T, W) \leq \frac{|T| + |W|}{(1 + \widetilde{M}^{-1})}.$$
Real Fourier Basis

Domain \( t = 0, 1, \ldots, N - 1 \), with \( N \) even,

\[
\varphi_0(t) = \frac{1}{\sqrt{N}},
\]

\[
\varphi_{2k-1}(t) = \sqrt{\frac{2}{N}} \sin\left(\frac{2\pi kt}{N}\right), \quad k = 1, 2, \ldots, N/2 - 1,
\]

\[
\varphi_{2k}(t) = \sqrt{\frac{2}{N}} \cos\left(\frac{2\pi kt}{N}\right), \quad k = 1, 2, \ldots, N/2 - 1,
\]

\[
\varphi_{N-1}(t) = \sqrt{\frac{1}{N}}(-1)^t,
\]

we have

\[
\tilde{M} = \sqrt{\frac{2}{N}},
\]

Modifying Earlier Arguments,

\[
\tilde{\mu}(T, W) < \sqrt{2(\|T\| + |W|)}/\sqrt{N}. \tag{13}
\]
Theorem 0.8 Let $\Phi_1$ be the basis of spikes and let $\Phi_2$ be the basis of real sinusoids. If $S$ is a superposition of atoms from sets $T$ and $W$ and

$$|T| + |W| \leq \frac{1}{2} \sqrt{N/2},$$

(14)

then

- the solution to $(P_0)$ is unique;
- the solutions of $(P_1)$ is unique;
- the two solutions are identical.
General Pairs of Bases

**Theorem 0.9** \( \Phi_1 \) and \( \Phi_2 \) orthonormal bases for \( \mathbb{R}^N \);

\[
M(\Phi_1, \Phi_2) = \sup\{ ||\langle \phi_1, \phi_2 \rangle || : \phi_1 \in \Phi_1, \phi_2 \in \Phi_2 \}.
\]

\( \Phi = \Phi_1 \cup \Phi_2 \) concatenation of the two bases.

Suppose \( S = \Phi \alpha \), where \( \alpha \) obeys

\[
||\alpha||_0 < \frac{1}{2} \left( 1 + M^{-1} \right),
\]

then \( \alpha \) is the unique solution to \( (P_1) \) and also the unique solution to \( (P_0) \).

Small \( M \) offers possibility of Ideal Atomic Decomposition.
Theorem 0.10  Let $\Phi_1$ and $\Phi_2$ be orthonormal bases for $\mathbb{R}^N$. Let $T$ index the collection of nonzero coefficients for $x$ in basis 1, and $W$ index the collection of nonzero coefficients for $x$ in basis 2. Then

$$|T| + |W| \geq (1 + M^{-1}).$$

(15)

- No signal can be analyzed in both bases and have simultaneously fewer than about $M^{-1}$ nonzero components from $\Phi_1$ and $\Phi_2$ together;

- A signal which is synthesized from fewer than about $M^{-1}$ components from $\Phi_1$ and $M^{-1}$ components from $\Phi_2$ is decomposed by minimum $\ell^1$ atomic decomposition perfectly into those components.
$M$ measures *mutual coherence* of two bases;

Small $M$ bases are *mutually incoherent*.

$0 \leq M \leq 1$; if two orthobases have an element in common, then $M = 1$.

**Lemma 0.11** *For any pair of orthonormal bases $\Phi_1, \Phi_2$ of $\mathbb{R}^N$,*

$$M(\Phi_1, \Phi_2) \geq 1/\sqrt{N}.$$
Random Orthogonal Bases

$\Phi_1, \Phi_2$ random orthogonal matrices, uniform on $O(N)$

$$M(\Phi_1, \Phi_2) \approx 2\sqrt{\log_e(N)}/\sqrt{N}$$

<table>
<thead>
<tr>
<th>Size $N$</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median $M(N)$</td>
<td>0.5684</td>
<td>0.4506</td>
<td>0.3543</td>
<td>0.2706</td>
<td>0.2052</td>
<td>0.1549</td>
</tr>
<tr>
<td>$2\sqrt{\log(N)/N}$</td>
<td>0.6582</td>
<td>0.5098</td>
<td>0.3894</td>
<td>0.2944</td>
<td>0.2208</td>
<td>0.1645</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.8636</td>
<td>0.8837</td>
<td>0.9099</td>
<td>0.9193</td>
<td>0.9296</td>
<td>0.9413</td>
</tr>
</tbody>
</table>

Table 1: Table of the medians of the maximum amplitude in a real $N \times N$ random orthogonal matrix, out of 100 generations.
Figure 5: Empirical distribution of the normalized maximum entry $M(N)/(2\sqrt{\log(N)/N})$ for $N = 64, 128, 256, 512$. Each is based on 1000 simulations.
Application: Speech Scrambling

A. D. Wyner (1979), Sloane (1983)

Scrambling real-valued discrete-time signal $S$.

- Form random orthogonal matrix $U$, known only to sender and intended recipient.

- Scrambling: $E = US$.

- Transmit $E$ in the clear,

- Intended Recipient $S = U^T E$. 
**Impulse Resistant Scrambling**

- Form random orthogonal matrix $U$, known only to sender and intended recipient.

- $M = \max_{i,j} |U_{ij}|$.

- Place $K < M^{-1}/2$ nonzero elements in $N$-vector $S$.

- Scrambling: $E = US$.

- Transmit $E$ in the clear.

- Intended Recipient: Minimum $\ell^1$ decomposition in dictionary consisting of (Spikes, Columns of U).
Corollary

Completely resistant to arbitrary changes in fewer than $K$ components of $E$.

In general: Transmit $O(\sqrt{N/\log(N)})$ real numbers in a vector of length $N$.

Immune to $O(\sqrt{N/\log(N)})$ gross errors.
MultiScale Bases

Key Feature of (Spike, Sinusoid) pair:

$M$ small for large $N$:

$$M = O(N^{-1/2})$$

More generally, $M$ roughly 1.

Orthobases $\Phi_1$ and $\Phi_2$; consider capacity defined by optimization problem

$$(K_\gamma) \quad \min \quad \|\Phi_1^T x\|_1 + \|\Phi_2^T x\|_1,$$

subject to $\langle x, \phi_\gamma \rangle = 1,$

Previous analysis $\text{Val}(K_\gamma)$ does not depend on $\gamma$, or at most weakly so.
Consider $\Phi_1$ Fourier, $\Phi_2$ Wavelets:

- $(K_{\gamma})$ near 1 at low frequencies
- $(K_{\gamma})$ small at high frequencies
Definition 0.12 Joint block diagonal structure

- **Orthogonal direct sum decomposition of** \( \mathbb{R}^N \) **as**

\[
\mathbb{R}^N = X_0 \oplus X_1 \oplus \ldots \oplus X_J.
\]

- **Grouping of indices** \( \Gamma_{1,j} \) **for basis 1 so that**

\[
\text{SPAN}(\phi_\gamma : \gamma \in \Gamma_{1,j}) = X_j
\]

**and grouping of indices** \( \Gamma_{2,j} \) **for basis 2 so that**

\[
\text{SPAN}(\phi_\gamma : \gamma \in \Gamma_{2,j}) = X_j
\]
Lemma 0.13  If basis pair joint block diagonal structure, then $(P_0)$ and $(P_1)$ direct sum of subproblems.

$S^{(j)}$ be the ortho-projection of $S$ on $X_j$,

$\Phi^{(j)}$ be the subdictionary formed from $\phi_\gamma$ with

$\gamma \in \Gamma_{1,j} \cup \Gamma_{2,j}$

Define

$(P_{0,j}) \quad \min \|\alpha^{(j)}\|_0, \quad \text{subject to} \quad S^{(j)} = \Phi^{(j)} \alpha^{(j)},$

and

$(P_{1,j}) \quad \min \|\alpha^{(j)}\|_1, \quad \text{subject to} \quad S^{(j)} = \Phi^{(j)} \alpha^{(j)}.$

If unique solution to each $(P_{0,j})$, solution to $(P_0)$ is concatenation of individual component solutions.

If a unique solution to each $(P_{1,j})$, solution to $(P_1)$ is concatenation of individual component solutions.
Lemma 0.14

\[ M_j = M(\{\phi_\gamma : \gamma \in \Gamma_{1,j}\}, \{\phi_\gamma : \gamma \in \Gamma_{2,j}\}) \]

**blockwise mutual incoherence.** Then if \( S \) is superposition of \( N_{1,j} \) terms from \( \Gamma_{1,j} \) and \( N_{2,j} \) terms from \( \Gamma_{2,j} \), and

\[ N_{1,j} + N_{2,j} < \frac{1}{2}M_j^{-1} \]

the solutions of each \((P_{0,j})\) and each \((P_{1,j})\) are unique and are the same.
Real bi-sinusoids $e_w$ derive Meyer-Lemarié wavelet construction.

With $\omega = (w, \sigma)$, where $w \in [2^j, 2^{j+1})$ and $\sigma \in \{1, 2\}$ we define $\Omega_j = [2^j, 2^{j+1}) \times \{1, 2\}$ and we have basis functions in four different groups:

**RW1.** $e_\omega(t) = b_j(w) \cos \left(2\pi w t / N\right) - b_j(w') \cos \left(2\pi w' t / N\right) \quad w < 2^j \cdot 4/3, \sigma = 1$;

**RW2.** $e_\omega(t) = b_j(w) \cos \left(2\pi w t / N\right) + b_j(w') \cos \left(2\pi w' t / N\right) \quad w \geq 2^j \cdot 4/3, \sigma = 1$;

**IW1.** $e_\omega(t) = b_j(w) \sin \left(2\pi w t / N\right) - b_j(w') \sin \left(2\pi w' t / N\right) \quad w < 2^j \cdot 4/3, \sigma = 2$;

**IW2.** $e_\omega(t) = b_j(w) \sin \left(2\pi w t / N\right) + b_j(w') \sin \left(2\pi w' t / N\right) \quad w \geq 2^j \cdot 4/3, \sigma = 2$.

Here $w'$ is the “twin” of $w$, and obeys

$$2^j - w' = w - 2^j, \quad w \leq 2^j \cdot 4/3;$$

$$2^{j+1} - w = w' - 2^{j+1}, \quad w > 2^j \cdot 4/3.$$
while—important point—$b_j(w)$ is a certain “bell function” that is also used in the construction of the Meyer Wavelet basis, and obeying

$$b_j(w)^2 + b_j(w')^2 = 2/N, \quad w \in [2^j, 2^{j+1}).$$

$(e_\omega)$ is orthonormal  
$(e_\omega)$ spans the same space $W_j$ as the collection of periodized Meyer wavelets.
**Application: Wavelets and bi-sinusoids**

\( \Phi_1 \) Meyer-Lemarié Wavelets

\( \Phi_2 \) real bi-sinusoids

**Lemma 0.15** *The wavelet coefficients at a given level \( j > j_0 \) are obtained from the real bi-sinusoid coefficients at the same level \( j \) by a finite orthogonal transform \( U_j \) of length \( 2^j \) built from discrete cosine and sine transforms.*

- The dictionary \( \Phi \) is joint block diagonal.

- \( M_j = 2^{-j/2} \)
To do ...

- More about Multiscale Bases
- More about Non-orthogonal Bases
- $M$ is unnecessarily pessimistic
An Experiment

The $\sqrt{N/\log(N)}$ is pessimistic...

Figure 6: Plot of median($\tilde{\mu}$) versus density $\nu$ for $N = 32, 64, 128$. The curve associated with $N = 32$ is the lowest and the curve associated with $N = 128$ is the highest.