Uncertainty Principles and Ideal Atomic Decomposition

David L. Donoho, *Member, IEEE,* and Xiaoming Huo

***Abstract*—Suppose a discrete-time signal $S(t)$, $0 \leq t < N$, is a superposition of atoms taken from a combined time–frequency dictionary made of spike sequences $1_{\{t=\tau\}}$ and sinusoids $\exp\{2\pi iwt/N\}/\sqrt{N}$. Can one recover, from knowledge of S alone, the precise collection of atoms going to make up S? Because every discrete-time signal can be represented as a superposition of spikes alone, or as a superposition of sinusoids alone, there is no unique way of writing S as a sum of spikes *and* sinusoids in general.**

We prove that if S is representable as a *highly sparse* superposition of atoms from this time–frequency dictionary, then there is only one such highly sparse representation of S, and it can be obtained by solving the *convex* optimization problem of minimizing the ℓ^1 norm of the coefficients among all decompositions. Here "highly sparse" means that $N_t + N_w < \sqrt{N}/2$ where N_t is the number of time atoms, N_w is the number of frequency atoms, and N is the length of the discrete-time signal.

Underlying this result is a general ℓ^1 uncertainty principle which says that if two bases are mutually incoherent, no nonzero signal can have a sparse representation in both bases simultaneously. For the above setting, the bases are sinuosids and spikes, and mutual incoherence is measured in terms of the largest inner product between different basis elements. The uncertainty principle holds for a variety of interesting basis pairs, not just sinusoids and spikes. The results have idealized applications to band-limited approximation with gross errors, to error-correcting encryption, and to separation of uncoordinated sources.

Related phenomena hold for functions of a real variable, with basis pairs such as sinusoids and wavelets, and for functions of two variables, with basis pairs such as wavelets and ridgelets. In these settings, if a function f is representable by a sufficiently sparse superposition of terms taken from both bases, then there is only one such sparse representation; it may be obtained by minimum ℓ^1 norm atomic decomposition. The condition "sufficiently sparse" becomes a multiscale condition; for example, that the number of wavelets at level j plus the number of sinusoids in the jth dyadic frequency band are together less than a constant times $2^{j/2}$.

***Index Terms*—Basis pursuit, combinatorial optimization, convex optimization, error-correcting encryption, harmonic analysis, Logan's phenomenon, matching pursuit, multiple-basis signal representation, overcomplete representation, ridgelet analysis, uncertainty principle, wavelet analysis.**

Manuscript received June 1, 1999; revised July 19, 2001. This work was supported by National Science Foundation under Grant DMS 95-05151, and by NSF under Grant ECS-97-07111.

D. L. Donoho is with the Statistics Department, Stanford University, Stanford, CA 94305 USA (e-mail: donoho@stanford.edu).

X. Huo is with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: xiaoming@isye.gatech.edu).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(01)08946-5.

I. Introduction

RECENTLY, workers in the computational harmonic analysis community have developed a number of interesting new signal representations; see [9], [23], [30]. In addition to sinusoids and wavelets, we now have Wilson bases, [10], wavelet packets, and cosine packets [8]. Moreover, the list of such representations is expanding all the time; recent additions include ridgelets and curvelets [5], [4].

In each of these cases, we have an orthonormal basis or tight frame which has been designed to be effective at representing objects of a specific type, where "effective" means requiring very few significant coefficients. The transforms turn out to be complementary in the sense that the type of objects for which one transform is well-suited are unlike the objects for which another transform is well-suited. For example, wavelets perform relatively poorly on high-frequency sinusoids, for which sinusoids are (naturally) very effective. On the other hand, sinusoids perform poorly on impulsive events, for which wavelets are very effective. In dimension 2, wavelets do poorly with discontinuities on edges, for which ridgelets are effective, while ridgelets do poorly on impulsive events [5].

It is natural in such a setting to imagine combining signal representations, using terms from each of several different bases. One supposes that the object of interest is a superposition of two phenomena, one of which by itself can be effectively represented in basis 1 and the other of which by itself can be effectively represented in basis 2, and one hopes that by allowing a representation built from terms in both bases, one might obtain an effective representation—far more effective than what one could obtain using either basis alone. Specifically, one hopes to represent an object containing two phenomena in superposition with the efficiency one would expect in analyzing each phenomenon separately in its own appropriate basis.

Such speculation leads one to propose the use of dictionaries $\Phi = \Phi_1 \cup \Phi_2 \cup \cdots \cup \Phi_D$ made from a concatenation of several orthonormal bases $\Phi_d = \{\varphi_{d,i}\}$, and to seek representations of a signal $S(t)$ as

$$S = \sum_{\gamma} \alpha_\gamma \varphi_\gamma \tag{1.1}$$

where $\gamma = (d, i)$ is an index into the dictionary, naming both the basis and the specific basis element. The general aim is to find concrete methods which offer decompositions of better sparsity through the use of several representations than is possible through any one representation alone. In the field of computational harmonic analysis, Mallat and Zhang [24] were early advocates of this approach, and introduced the "dictionary

methodology." Coifman and collaborators have made numerous related contributions in this field also. In the field of signal processing, a considerable body of related practical work has been done in the general area of multiple-basis signal representation and signal compression. See, for example, the survey paper of Berg and Mikhael [3], who mention seminal work of a decade ago by Mikhael and Spanias [26] and Beex [2]. There is ongoing more recent work by DeBrunner and collaborators in signal representation [12] and in image representation [11]. Finally, we mention the thesis of Huo [19].

Mallat and Zhang [24] proposed a heuristic greedy approximation method for representation using overcomplete dictionaries, called Matching Pursuit. While Matching Pursuit works well in many cases, it is known not to provide sparse approximations in general; see the counterexamples in [7], [13].

As Φ is the concatenation of several bases, the representation (1.1) is not unique; any single basis alone affords already decomposition of an arbitrary signal S, and consequently many possibilities for combined decomposition arise. The general goal would be to find a highly sparse decomposition—one with very few nonzero terms. This leads to the optimization problem

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

where $\|\alpha\|_0 = \#\{\gamma\colon \alpha_\gamma \neq 0\}$ is the ℓ^0 quasi-norm. Unfortunately, in general, this problem requires a search through subsets of Φ looking for a sparse subset providing exact decomposition.

Chen, Donoho, and Saunders [6], [7] proposed an alternate approach to signal decomposition in dictionaries, which they called Basis Pursuit (BP). It calls for solving the ℓ^1 optimization problem

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

where $\|\alpha\|_1 = \sum |\alpha_\gamma|$ is the ℓ^1 norm of the coefficients. This is a convex optimization problem, and can be attacked using linear programming methods based either on the classical simplex method of linear programming or the recently popular interior point methods [32]. As the ℓ^1 norm is, in a certain natural sense, a convexification of the ℓ^0 norm, the problem (P_1) can be viewed as a convexification of (P_0), one which makes accessible a variety of computationally feasible strategies.

In [7], it was shown that, empirically, the solution of BP is frequently quite sparse; and that in fact when the underlying synthesis was made from only a few dictionary elements, the BP solution may *perfectly recover* the specific atoms and specific coefficients used in the synthesis. For example, Chen considered a sum of four sinusoids and two spikes, decomposed them in a combined time–frequency dictionary of sinusoids and spikes, and found that BP recovered exactly the indexes and coefficients of the terms involved in the synthesis; this held across a wide range of amplitude ratios between the sinusoid and spike components. In contrast, the same signal was analyzed using Matching Pursuit, the recovery of indexes and coefficients was only approximate and became very inexact when the sinusoidal and spike components were at very different amplitudes.

A. Ideal Atomic Decomposition

Our goal in this paper is to prove that in certain specific cases, when the signal is a sufficiently sparse sum of terms from a dictionary, the BP principle of ℓ^1 optimization of the decomposition from that dictionary in fact gives the solution of the ℓ^0 optimization problem and in fact recovers the identities and coefficients of the original synthesizing elements perfectly.

The following terminology helps formalize this phenomenon. If Φ is an overcomplete system, any representation $S = \sum_\gamma \alpha_\gamma \phi_\gamma$ is an *atomic decomposition* using *atoms from the dictionary*. If S in fact can be generated by a highly sparse sum, with the term "highly sparse" given an appropriate definition, and there is in fact only one such highly sparse way of doing so, and if an optimization principle finds that decomposition, we say that the principle leads to *ideal atomic decomposition* under the stated sparsity hypothesis. In effect then, we are claiming that under certain sparsity conditions, the minimum ℓ^1-norm decomposition in certain dictionaries achieves an ideal atomic decomposition.

B. Time–Frequency Decomposition

We initially consider the situation where $\Phi = \Phi_1 \cup \Phi_2$ with Φ_1 the spike basis

$$\varphi_{1,\tau}(t) = 1_{\{t=\tau\}}, \qquad \tau = 0, 1, \ldots, N-1$$

and Φ_2 the Fourier basis

$$\varphi_{2,w}(t) = \frac{1}{\sqrt{N}} \exp(2\pi i w t/N), \qquad w = 0, 1, \ldots, N-1.$$

Both Φ_1 and Φ_2 are orthonormal bases for l_N^2.

Theorem I.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 nonunique solution.

Theorem I.2: Let

$$S = \sum_{\gamma \in T \cup W} \alpha_\gamma \varphi_\gamma$$

with T, W as in Theorem I.1. If

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

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 nonunique solution.

In short, if the signal S truly has a very sparse decomposition in the time–frequency dictionary, this is unique, and basis pursuit (ℓ^1 decomposition) will find it.

C. Relation to the Uncertainty Principle

Underlying Theorems I.1 and I.2 is an uncertainty principle: the *analysis* of a signal in the time and frequency domains

cannot yield a transform pair which is sparse in both domains simultaneously.

To explain this connection, note that in order to take ideal atomic decomposition seriously we must know that 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*. If this were possible, the atomic decomposition would be nonunique.

Now suppose there existed a signal whose Fourier transform was very sparse *and* whose representation in the standard basis was very sparse. Then we would have exactly an example of such nonunique sparse decomposition: the signal could be represented in two different ways: as a sparse sum of sinusoids and as a sparse sum of spikes.

In effect, at the center of our analysis of the ℓ^1 decomposition in this finite-N, discrete-time setting is exactly a certain picket fence sequence $\mathbf{III}$ which may equally be viewed either as a relatively sparse sum of sinusoids or an equally sparse sum of spikes. This sequence has been studied before in connection with the uncertainty principle, for which it serves as a kind of extremal function [15]; see also recent work of Przebinda, DeBrunner, and Özaydin [27].

The connection between unique decomposition and the uncertainty principle will emerge repeatedly, and in a quantitative form, throughout the paper. It is closely connected to work on the uncertainty principle in [15], [16], however, the uncertainty principle employed here gives a more symmetric role for time and frequency.

D. Nonlinearity of the ℓ^1 Norm

The phenomenon of ideal atomic decomposition is very special; it follows from very particular properties of the ℓ^1 norm. In effect, (P_1) asks to find the member of a linear subspace closest to the origin in ℓ^1 norm. This closest point problem (which would be a linear problem in ℓ^2 norm) is highly nonlinear in ℓ^1 norm, and the nonlinearity is responsible for our phenomenon.

A precedent for this type of perfect recovery is what has been called Logan's Phenomenon in [15]; see also [22], [16]. That phenomenon occurs when one is trying to find a decomposition of a signal into band-limited function and impulsive noise; supposing that the product of the signal bandwidth and the measure of the support of the noise is sufficiently small, this can be done perfectly, by finding the band-limited function closest to the observed signal in an ℓ^1 sense. The phenomenon is highly nonlinear in the sense that perfect reconstruction holds at all signal/noise ratios. See Section V below.

In a sense, the phenomenon exposed in this paper is due to the same nonlinearity of the ℓ^1 norm, only transposed into the setting of approximation from arbitrary time–frequency dictionaries in which time and frequency play a symmetric role, and in which there is no need for the frequency support of the signal to be an interval or even to be known.

E. Other Dictionary Pairs

In fact, the methods of this paper provide insights outside of the setting of time–frequency pairs. We give two examples. The first considers dictionaries of sinusoids and wavelets.

Theorem I.3: Let $f(\theta)$ denote a square-integrable function on the circle $[0, 2\pi)$. Suppose that f is a superposition of sinusoids and wavelets

$$f(\theta) = \sum_{\lambda} \alpha_\lambda \psi_\lambda(\theta) + \sum_{n \geq n_0}^{\infty} c_n e^{in\theta}. \tag{1.2}$$

Here, the ψ_λ 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

$$N_j(\text{WAVELETS}) + N_j(\text{SINUSOIDS}) \leq C \cdot 2^{j/2}, \qquad j = j_0 + 1, \ldots . \tag{1.3}$$

Consider the overcomplete dictionary Φ consisting of Meyer–Lemarié wavelets and of high-frequency sinusoids $e^{in\theta}$, $n \geq n_0 \geq 2^{j_0+1}$. There is at most one way of decomposing a function f in the form (1.2) while obeying (1.3). If f has such a decomposition, it is the unique solution to the minimum ℓ^1 optimization problem

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

In short, minimum ℓ^1 decomposition, which makes no assumption about the sparsity or nonsparsity of the representation of f, nevertheless gives ideal atomic decomposition when sufficient sparsity is present.

Note however, that the notion of sparsity becomes *level-dependent*. We can tolerate more total terms at high resolution than we can at low resolution. Intuitively, this is because there is less possibility of confusion between sparse sums of wavelets and sparse sums of sinusoids as we go to sums limited to dyadic bands at increasingly high frequencies—the two systems become increasingly disjoint.

Mathematically, we could say that there is an uncertainty principle: a phenomenon near scale 2^{-j} and frequency 2^j cannot have a sparse representation in both the wavelets basis and the sinusoid basis. The expression of this phenomenon is the fact that if a function f has at most $C \cdot 2^{j/2}$ nonzero wavelet coefficients and sinusoid coefficients at level j, then the function is zero.

For a second example of this kind, we consider combined dictionaries of wavelets and ridgelets.

Theorem I.4: Let $f(x_1, x_2)$ denote a square-integrable function on $\boldsymbol{R}^2$. Suppose that f is a superposition of wavelets and ridgelets

$$f = \sum_{Q} \alpha_Q \psi_Q + \sum_{\lambda \in \Lambda}^{\infty} \beta_\lambda \rho_\lambda. \tag{1.4}$$

Here, the ψ_Q are the usual two-dimensional (2-D) Meyer–Lemarié wavelets for the plane. The ρ_λ are orthonormal ridgelets [14] and Λ consists of ridgelets at ridge scales

$j \geq j_0 + 2$. There is a constant $C > 0$ with the following property. Let $N_j(\text{WAVELETS})$ be the number of wavelets used in this decomposition at resolution level j and let $N_j(\text{RIDGELETS})$ be the number of ridgelets at level j. Suppose that the sum obeys all the conditions

$$N_j(\text{WAVELETS}) + N_j(\text{RIDGELETS}) \leq C \cdot 2^{j/2}, \quad j = j_0 + 2, \ldots. \quad (1.5)$$

Consider the overcomplete dictionary Φ consisting of Meyer–Lemarié wavelets and of ridgelets with $\lambda \in \Lambda$. There is at most one way of decomposing a function f in the form (1.4) while obeying (1.5). If f has such a decomposition it is the unique solution of the minimum ℓ^1 optimization problem

$$\sum_Q |\alpha_Q| + \sum_\lambda |\beta_\lambda|.$$

In short, minimum ℓ^1 decomposition, which makes no assumption about the sparsity or nonsparsity of the representation of f, nevertheless gives ideal atomic decomposition when sufficient sparsity is present.

Again, the notion of sparsity becomes *level-dependent*. We again tolerate more total terms at high resolution than we do at low resolution. Intuitively, this is because there is less possibility of confusion between sparse sums of wavelets and sparse sums of ridgelets as we go to sums limited to dyadic bands at increasingly high frequencies—the two systems become increasingly disjoint.

Mathematically, we could say that there is an uncertainty principle: a phenomenon occurring at scale 2^{-j} and frequency 2^j cannot have a sparse representation in both the wavelets basis and the ridgelets basis. The expression of this phenomenon is the fact that if a function f has at most $C \cdot 2^{j/2}$ nonzero wavelet coefficients and ridgelet coefficients at level j, then the function is zero.

F. Contents

Sections II–IV of the paper prove Theorems I.1 and I.2. Section V gives an application to band-limited approximation with unknown band and impulsive noise. Section VI discusses generalizations of Theorems I.1 and I.2 to the setting of real sinusoids (as opposed to complex exponentials). Section VII isolates the concept—mutual incoherence—which makes Theorems I.1 and I.2 work, and shows that it generalizes to other pairs of orthogonal bases; Section VIII shows that in some sense "most pairs of ortho bases" are mutually incoherent. It also gives applications to encryption and blind separation of uncoordinated sources. Sections IX–XI switch gears, establishing Theorems I.3 and I.4. Section XII describes generalizations to the nonorthogonal setting. Section XIII considers relations of the concepts here to the classical uncertainty principle for functions of a single real variable, and applies insights derivable from experience in that setting. It also suggests that for many situations, the provable bound $|T| + |W| < \text{const} \cdot \sqrt{N}$ of Theorems I.1 and I.2 overstates severely the required sparsity; often $|T| + |W| < \text{const} \cdot N$ is sufficient for uniqueness.

II. Uniqueness of ℓ^0 Optimization

We begin by quoting a simple uncertainty principle from [15].

Theorem II.1: 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

$$N_t + N_w \geq 2\sqrt{N}. \quad (2.1)$$

The proof identifies the extremal functions for these inequalities. When N is a perfect square, (2.1) is attained by

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

and by its frequency and time shifts. The complete catalog of extremal functions is generated by scalar multiples of

$$\exp\{2\pi i/N \cdot w \cdot (t \ominus \tau)\}\mathbf{III}_{t \ominus \tau}$$

where w is an integer in the range $0 \leq w < \sqrt{N}$, τ is an integer in the range $0 \leq \tau < \sqrt{N}$, and $\ominus$ denotes subtraction modulo N.

The key properties of $\mathbf{III}$ are its sparsity $(N_t + N_w = 2\sqrt{N})$ and its invariance under Fourier transformation

$$\mathcal{F}(\mathbf{III}) = \mathbf{III}.$$

This says that $\mathbf{III}$ may equally well be viewed as either being produced by

1) time-domain synthesis using $\sqrt{N}$ spikes, or
2) frequency-domain synthesis from $\sqrt{N}$ sinusoids.

In consequence, for $S = \mathbf{III}$, the problem (P_0) has a nonunique solution in the overcomplete dictionary

$$\{\text{SPIKES}\} \cup \{\text{SINUSOIDS}\}.$$

It follows that constraints on sparsity $N_t + N_w < K$ cannot guarantee uniqueness in this setting for $K > \sqrt{N}$. In fact, $K = \sqrt{N}$ can guarantee uniqueness, as we have claimed previously in Theorem I.1. We now show this, and thereby prove Theorem I.1.

Suppose that S had two decompositions $S = \Phi\alpha^1$, $S = \Phi\alpha^2$, where both α^1 and α^2 obey $\|\alpha^i\|_0 < \sqrt{N}$; then $0 = \Phi(\alpha^1 - \alpha^2)$. In other words, if we let $\mathcal{N}$ denote the nullspace $\mathcal{N} = \{\delta : \Phi\delta = 0\}$, then $\alpha^1 - \alpha^2 \in \mathcal{N}$. For $\delta \in \mathcal{N}$, suppose

$$\delta = (\delta_{(1,0)}, \delta_{(1,1)}, \ldots, \delta_{(1,N-1)}, \delta_{(2,0)}, \delta_{(2,1)}, \ldots, \delta_{(2,N-1)})$$

where the first N components are associated with dictionary elements belonging to the spike basis and the last N are associated with dictionary elements belonging to the Fourier basis. Thus, partitioning $\delta = (\delta^1, \delta^2)$ into components, $\delta \in \mathcal{N}$ implies

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

or

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

In a more transparent notation, $\mathcal{N}$ is the set of all pairs $(x, -\hat{x})$, where $x = (x_t)_{t=0}^{N-1}$ and $\hat{x} = (\hat{x}_w)_{w=0}^{N-1}$ is its Fourier transform.

Returning now to our setting, $\delta = \alpha^1 - \alpha^2$ has, therefore, the structure of a pair $(x, -\hat{x})$; by the uncertainty principle in Theorem II.1, δ must have at least $2\sqrt{N}$ nonzero entries or else

$\delta = 0$. But by hypothesis $\|\alpha^1\|_0 < \sqrt{N}$ and $\|\alpha^2\|_0 < \sqrt{N}$. Hence, $\delta = 0$; in short, $\alpha^1 = \alpha^2$.

III. Uniqueness of ℓ^1 Optimization

Suppose that $S = \Phi\alpha$, where α is sparse, made from atoms in sets T and W in the time and frequency domain, respectively. We seek a condition on the size of T and W which guarantees that α is the unique solution of the ℓ^1 optimization problem (P_1).

In order that α be the unique solution, we must have $\|\tilde{\alpha}\|_1 > \|\alpha\|_1$, for every $\tilde{\alpha}$ satisfying $\Phi\tilde{\alpha} = \Phi\alpha$. Equivalently, for every $\delta \in \mathcal{N}$ $(\Phi\delta = 0)$, we must have

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

unless $\delta = 0$. Now

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

Note that

$$|\alpha_\gamma + \delta_\gamma| - |\alpha_\gamma| \geq -|\delta_\gamma|$$

and so

$$\|\alpha + \delta\|_1 - \|\alpha\|_1 \geq \sum_{(T\cup W)^c} |\delta_\gamma| - \sum_{T\cup W} |\delta_\gamma|.$$

Hence, a sufficient condition for uniqueness is that for $\delta \neq 0$

$$\sum_{T\cup W} |\delta_\gamma| < \sum_{(T\cup W)^c} |\delta_\gamma|, \qquad \forall\, \delta \in \mathcal{N}. \tag{3.1}$$

In words, every nonzero member of $\mathcal{N}$ has a smaller ℓ^1 norm on the support of α than off the support of α. Since $\mathcal{N}$ consists of all pairs $(x, -\hat{x})$, the condition (3.1) is equivalent to

$$\sum_T |x_t| + \sum_W |\hat{x}_w| < \tfrac{1}{2}(\|x\|_1 + \|\hat{x}\|_1) \tag{3.2}$$

for every nonzero x. Formalizing matters somewhat, we view this as a time-frequency concentration problem. For given sets T and W, let

$$\mu(T, W) = \sup \frac{\sum_T |x_t| + \sum_W |\hat{x}_w|}{\|x\|_1 + \|\hat{x}\|_1} \tag{3.3}$$

where the supremum is over all $x = (x_t)_{t=0}^{N-1}$ which are nonzero. This measures the degree to which the joint ℓ^1 norm can be concentrated to sets T and W; the uniqueness of ℓ^1 optimization is therefore implied by

$$\mu(T, W) < \tfrac{1}{2}. \tag{3.4}$$

We note that $\mu(T, W)$ is closely related to a variety of known time-frequency concentration functionals connected with the uncertainty principle. See Section V.

The sequence $\mathbf{III}_t$ shows that we can have

$$\mu(T, W) \geq \frac{|T| + |W|}{2\sqrt{N}} \tag{3.5}$$

and, in particular, if $\sqrt{N}$ is even, there exist $|T|$, $|W|$ of size $\frac{\sqrt{N}}{2}$ so that

$$\mu(T, W) \geq \frac{1}{2}. \tag{3.6}$$

In short, for a sparsity condition on T and W to imply uniqueness of a solution to (P_1), it must clearly be of the form $|T| + |W| < K$, for some $K \leq \sqrt{N}$. This is the same range as we contemplated in the condition for uniqueness in the ℓ^0 problem (P_0), but it is a necessary restriction: we can see from the sequence $\mathbf{III}$ that there are sets $|T| = \sqrt{N}$ so that the problem (P_1) has a nonunique solution. Indeed

$$\mathbf{III} = \Phi_1 \cdot \mathbf{III} = \Phi_2 \cdot \widehat{\mathbf{III}}$$

and one can verify that

$$\alpha^1 = \begin{cases} 1, & \gamma = (1, t),\ t \in \operatorname{supp}\{\mathbf{III}\} \\ 0, & \text{else} \end{cases}$$

$$\alpha^2 = \begin{cases} 1, & \gamma = (2, w),\ w \in \operatorname{supp}\{\widehat{\mathbf{III}}\} \\ 0, & \text{else} \end{cases}$$

are both solutions of the problem (P_1), as are all convex combinations of α^1 and α^2. Curiously, $\mathbf{III}$ is within a factor 2 extremal for the $\mu(T, W)$ concentration measure.

Theorem III.1: Let T be a subset of the time domain and W be a subset of the frequency domain. Then

$$\mu(T, W) \leq \frac{|T| + |W|}{\sqrt{N} + 1}. \tag{3.7}$$

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.

We need two lemmas.

Lemma III.2: Let $(x, \hat{x})$ be a Fourier transform pair. Then

$$\|\hat{x}\|_1 \geq \sqrt{N}\|x\|_\infty. \tag{3.8}$$

Proof: Let $|x_t| = \|x\|_\infty$. Then from Fourier inversion

$$x_t = \sum_{w=0}^{N-1} \hat{x}_w e_w(t)$$

with e_w the l^2-normalized sinusoid of frequency $\frac{2\pi w}{N}$, so that

$$|x_t| \leq \|\hat{x}\|_1 \|e_w\|_\infty.$$

Now

$$\|e_w\|_\infty = \max_t \left|\frac{1}{\sqrt{N}} \exp\{2\pi i w t/N\}\right| = \frac{1}{\sqrt{N}}.$$

Equation (3.8) follows. □

Lemma III.3: Consider the capacity defined by the optimization problem

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

The value of this optimization problem obeys

$$\operatorname{Val}(K_{1,\tau}) = \operatorname{Val}(K_{1,0}), \qquad \tau = 1, 2, \ldots, N-1. \tag{3.9}$$

Also, for the frequency-side capacity defined by the optimization problem

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

we have

$$\text{Val}(K_{2,w}) = \text{Val}(K_{2,0}), \qquad w = 1, 2, \ldots, N-1. \quad (3.10)$$

Finally

$$\text{Val}(K_{1,0}) = \text{Val}(K_{2,0}) = 1 + \sqrt{N}. \qquad (3.11)$$

Proof: Candidates $x^{(1,0)}$ for $(K_{1,0})$ and $x^{(1,\tau)}$ for $(K_{1,\tau})$ are related by appropriate translation–modulation

$$x_t^{(1,\tau)} = x_{t \ominus \tau}^{(1,0)}$$
$$\hat{x}_w^{(1,\tau)} = \exp\{i2\pi w\tau\}\hat{x}_w^{(1,0)}.$$

Indeed, this transformation preserves the l^1 norm $||x||_1 + ||\hat{x}||_1$ and the constraint $x_0 = 1$ maps to $x_\tau = 1$. Hence, any solution of $(K_{1,0})$ maps to a solution of $(K_{1,\tau})$, and *vice versa*. Similar ideas map solutions of $(K_{2,w})$ into solutions of $(K_{2,0})$, and *vice versa*.

Similarly, the formal interchange of time and frequency domains turns any candidate for $(K_{1,0})$ into a candidate for $(K_{2,0})$ with equal constraint and equal norm. Finally, from Lemma III.2, we have

$$\text{Val}(K_{1,0}) \geq 1 + \sqrt{N}.$$

On the other hand, let $x = \varphi_{(1,0)}$ be the Kronecker sequence. Then x obeys the constraint of $(K_{1,0})$ while

$$||x||_1 + ||\hat{x}||_1 = 1 + \sqrt{N}.$$

Equation (3.11) follows. □

The proof of Theorem III.1 follows directly from Lemma III.3

$$|x_t| \leq \text{Val}(K_{1,t})^{-1} \left(||x||_1 + ||\hat{x}||_1\right)$$
$$|\hat{x}_w| \leq \text{Val}(K_{2,w})^{-1} \left(||x||_1 + ||\hat{x}||_1\right)$$

so

$$\frac{\sum_T |x_t| + \sum_W |\hat{x}_w|}{||x||_1 + ||\hat{x}||_1} \leq \sum_T \text{Val}(K_{1,t})^{-1} + \sum_W \text{Val}(K_{2,w})^{-1}$$
$$= (|T| + |W|)\left(\sqrt{N} + 1\right)^{-1}.$$

Equation (3.7) follows.

IV. Simultaneous Solution of ℓ^0 and ℓ^1

From the results of Section III, we know that a solution to (P_1), if it satisfies $||\alpha||_0 < \frac{1}{2}\sqrt{N}$, is unique. This must also solve (P_0), because at most one vector α may satisfy $S = \Phi\alpha$ and $||\alpha||_0 < \frac{1}{2}\sqrt{N}$. In short, any vector α obeying $||\alpha||_0 < \frac{1}{2}\sqrt{N}$ and $S = \Phi\alpha$ is simultaneously the solution of (P_1) and (P_0).

V. Application: Band-Limited Approximation With Gross Errors

Before continuing with our development of general atomic decomposition results, we indicate an application. The functional $\mu(T, W)$ we have studied in Section III is related to time-frequency concentration functionals connected with band-limited approximation. Donoho and Stark [15] defined

$$\mu_0(T, W) = \sup \frac{\sum_T |x_t|}{||x||_1}, \qquad \text{subject to } \text{supp}\{\hat{x}_w\} \subset W.$$

In short, this measures the time-side concentration ratio $\sum_T |x_t| / \sum_{t=0}^{N-1} |x_t|$ for objects x perfectly localized to W on the frequency side. They gave the inequality

$$\mu_0(T, W) \leq |T||W|/N$$

and described applications in the recovery of band-limited signals facing scattered gross errors. They assumed that one observed

$$S(t) = B(t) + \epsilon(t)$$

where B is a discrete-time band-limited signal with frequency-domain support purely in a certain *known* band W and that ϵ is a discrete-time noise, of arbitrary size, supported in a set T. In that setting they showed that whenever the support of the noise satisfies

$$\mu_0(T, W) < 1/2$$

the ℓ^1 approximant

$$\tilde{B} = \arg\min_X ||S - X||_1 \qquad \text{subject to } \text{supp}(\hat{X}) \in W$$

recovers B perfectly: $\tilde{B} = B$. Here, W is a *known* frequency band, but the support T of the noise is unknown. This is an instance of what they called *Logan's phenomenon* for band-limited ℓ^1 approximation, after B. F. Logan, who discovered it in the setting of low-pass approximation to continuous-time signals [22]. Compare also [16].

The concentration notion μ given in this paper is not directly comparable with μ_0, nor is the application of ℓ^1 approximation the same as ℓ^1 decomposition. In [15], the μ_0 functional supposes that the object in question is perfectly localized to a set W in the frequency domain, and measures the degree of concentration to T, while in this paper, the object is not assumed to be perfectly localized either to T or to W, and the quantity μ is fully symmetric in the roles played by time and frequency. Also, the ℓ^1 approximation in [15] was based on finding the ℓ^1-closest approximant from a fixed, known band W. In short, the signal was representable as a superposition of sinusoids with fixed and known frequencies. In contrast, the ℓ^1 decomposition here is based on approximation from an arbitrary collection of times and/or frequencies, none of which is prespecified. The method uses whatever combination of spikes and/or sinusoids may be necessary to decompose the object. If we label $\tilde{B}$ as the component of the ℓ^1 solution coming from sinusoids and $\tilde{\epsilon}$ as the component of the ℓ^1 solution coming from spikes, the approach of this paper may be viewed as a method for also solving the

problem of band-limited approximation with *unknown* band W! The results of this paper show that, if

$$|\mathrm{supp}(\hat{B})| + |\mathrm{supp}(\epsilon)| \le \frac{1}{2}\sqrt{N}$$

then $\tilde{B} = B$ and $\tilde{\epsilon} = \epsilon$.

In short, the ℓ^1 atomic decomposition may be viewed as a method for recovery of a band-limited signal with unknown band W in the presence of sparse gross errors in the time domain. The errors may be of arbitrary amplitude, but if the band W and the support T of the errors are both sufficiently sparse, then ℓ^1 atomic decomposition gives perfect reconstruction.

In comparing the approach of this paper with the older one, we see a key difference: namely, that the condition for perfect recovery in the band-limited approximation algorithm is $|T||W| < N/2$, whereas the condition in the atomic decomposition algorithm is $|T| + |W| < \frac{1}{2}\sqrt{N}$; the conditions cover a somewhat different collection of T, W pairs.

VI. Real Sinusoids

So far, we have been using as sinusoid basis the traditional system of complex exponentials $(\frac{1}{\sqrt{N}} e^{i2\pi wt/N})_{w=0}^{N-1}$. How do things change if we use instead the real sinusoids, or one of the discrete cosine transform or discrete sine transform bases [28]?

Let $(\varphi_w)_{w=0}^{N-1}$ be an orthonormal system for $l_{2,N}$. Let $\tilde{x}_w = \langle x, \varphi_w \rangle$ be the Fourier–Bessel coefficients in this system. Let T and W be subsets of the t- and w-index space, respectively. Define

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

so that what we earlier called $\mu(T, W)$ is the special case with $\varphi_w = \frac{1}{\sqrt{N}} e^{i2\pi wt/N}$. Careful inspection of previous arguments will show that if we put

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

then for problem

$$(\tilde{K}_{1,t}) \qquad \min \|x\|_1 + \|\tilde{x}\|_1, \qquad \text{subject to } x_t = 1$$

we have

$$\mathrm{Val}(\tilde{K}_{1,t}) \ge 1 + \tilde{M}^{-1}, \qquad \forall t$$

and, similarly, for problem

$$(\tilde{K}_{2,w}) \qquad \min \|x\|_1 + \|\hat{x}\|_1, \qquad \text{subject to } \hat{x}_w = 1$$

we have

$$\mathrm{Val}(\tilde{K}_{2,w}) \ge 1 + \tilde{M}^{-1}, \qquad \forall w.$$

It follows that

$$\tilde{\mu}(T, W) \le \frac{|T| + |W|}{(1 + \tilde{M}^{-1})}.$$

Now for the real Fourier basis, for domain $t = 0, 1, \ldots, N-1$, with N even

$$\varphi_0(t) = 1/\sqrt{N}$$

$$\varphi_{2k-1}(t) = \sqrt{\frac{2}{N}} \sin(2\pi kt/N), \qquad k = 1, 2, \ldots, N/2 - 1$$

$$\varphi_{2k}(t) = \sqrt{\frac{2}{N}} \cos(2\pi kt/N), \qquad k = 1, 2, \ldots, N/2 - 1$$

$$\varphi_{N-1}(t) = \sqrt{\frac{1}{N}} (-1)^t$$

we have

$$\tilde{M} = \sqrt{\frac{2}{N}}$$

and so

$$\tilde{\mu}(T, W) < \sqrt{2}(|T| + |W|)/\sqrt{N}. \tag{6.1}$$

Combining this with arguments of Section III, we immediately obtain the following.

Theorem VI.1: Let Φ_1 be the basis of spikes and let Φ_2 be the basis of real sinusoids. If S is a superposition of atoms from sets T and W and

$$|T| + |W| \le \tfrac{1}{2}\sqrt{N/2} \tag{6.2}$$

then the solution to (P_1) is unique.

What about the solution of (P_0)? Arguing as in Section II, we wish to ask about the minimal cardinality of sets T and W so that a pair $(x, \tilde{x})$ exists with x concentrated to T and $\tilde{x}$ concentrated to W. Unless $\tilde{\mu}(T, W) \ge 1$ there is no signal x 100% concentrated to T, with its real Fourier transform $\tilde{x}$ is also perfectly concentrated to W. The inequality (6.1) therefore shows that such perfect concentration is impossible, unless

$$|T| + |W| > \sqrt{N/2}.$$

We conclude as follows.

Theorem VI.2: Let Φ_1, Φ_2, S, T, and W be the same as in Theorem VI.1. If

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

then

- the solution to (P_0) is unique;
- the solutions of problem (P_0) and (P_1) are identical.

Actually, the criterion of uniqueness for the ℓ^0 problem can be sharpened by a factor two. The key is the following uncertainty principle for the real Fourier transform:

Theorem VI.3: Let x be the coefficient vector associated with the spike basis and let $\tilde{x}$ be the coefficient vector associated with the real Fourier basis. Suppose x and $\tilde{x}$ have N_t and $\tilde{N}_w$ nonzero elements, respectively. We have

$$N_t \cdot \tilde{N}_w \ge N/2$$

and so

$$N_t + \tilde{N}_w \ge \sqrt{2N}.$$

A variation of **III** will achieve $N_t + \tilde{N}_w = \sqrt{2N} + 1$.

Proof: Let $\hat{x}$ be the complex Fourier transform of x. The two sequences $\hat{x}$ and $\tilde{x}$ are connected in the following way: for an even N

$$\begin{aligned}\tilde{x}_0 &= \hat{x}_0 \\ \tilde{x}_{N-1} &= \hat{x}_{N/2} \\ \tilde{x}_{2k} &= \hat{x}_k + \hat{x}_{N-k}, \qquad k = 1, 2, \ldots, N/2-1 \\ \tilde{x}_{2k-1} &= i(\hat{x}_k - \hat{x}_{N-k}), \qquad k = 1, 2, \ldots, N/2-1.\end{aligned}$$

Letting N_w denote the number of nonzero elements in $\hat{x}$, we have $\tilde{N}_w \geq \frac{1}{2}N_w$. By Theorem II.1, $N_t \cdot N_w \geq N$, hence $N_t \cdot \tilde{N}_w \geq \frac{1}{2}N$.

Suppose $\sqrt{N/2}$ is an integer. As a variation of $\mathbf{\amalg}$, we consider

$$x_0 = x_{\sqrt{2N}} = x_{2\sqrt{2N}} = \cdots = x_{N-\sqrt{2N}} = 1$$

with other entries of x vanishing. Hence,

$$\hat{x}_0 = \hat{x}_{\sqrt{N/2}} = \hat{x}_{2\sqrt{N/2}} = \cdots = \hat{x}_{N-\sqrt{N/2}} = 1$$

with other entries of $\hat{x}$ vanishing. Then

$$\begin{aligned}\tilde{x}_0 &= 1, \\ \tilde{x}_{N-1} &= 1, \\ \tilde{x}_{\sqrt{2N}} &= \tilde{x}_{2\sqrt{2N}} = \tilde{x}_{3\sqrt{2N}} = \cdots = \tilde{x}_{N-\sqrt{2N}} = 1\end{aligned}$$

and the rest of $\tilde{x}$ vanishes. Hence, $N_t = \sqrt{N/2}$, $N_w = \sqrt{2N}$, and $\tilde{N}_w = \sqrt{N/2}+1$. □

The following theorem gives uniqueness of ℓ^0 optimization for a frequency domain based on the real Fourier transform.

Theorem VI.4: Let Φ_1, Φ_2, S, T, and W be the same as in Theorem VI.1. If

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

then the solution to (P_0) is unique. There are S, T, W with

$$|T| + |W| > \sqrt{N/2}$$

for which the solution to (P_0) is not unique.

In short, we have a parallel of the earlier situation based on the complex Fourier transform, only with a lower threshold for the $(P_1) \Leftrightarrow (P_0)$ equivalence effect. There is a similar parallel, with the same lower threshold, for the various real orthogonal bases associated with the real discrete cosine transforms and discrete sine transforms.

VII. Mutual Incoherence

The extension from complex sinusoids to real sinusoids generalizes immediately, to the following result.

Theorem VII.1: Let Φ_1 and Φ_2 be orthonormal bases for R^N and let

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

Let $\Phi = \Phi_1 \cup \Phi_2$ be the concatenation of the two bases. Let $S = \Phi\alpha$, where α obeys

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

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

This shows that sufficiently small values of the functional $M(\Phi_1, \Phi_2)$ guarantee the possibility of ideal atomic decomposition. We call M a measure of the *mutual coherence* of two bases; if two bases have a very small value of M then we say that they are mutually incoherent. Obviously $0 \leq M \leq 1$; if two orthobases have an element in common, then $M = 1$. On the other hand

$$M(\text{SPIKES, COMPLEX SINUSOIDS}) = 1/\sqrt{N}$$

so that M can be small for large N. There is an easy bound on how incoherent two bases can be.

Lemma VII.2: For any pair of orthonormal bases Φ_1, Φ_2 of $\boldsymbol{R}^N$

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

Proof: The matrix $\Phi_1^T\Phi_2$ is an orthonormal matrix. The sum of squares of entries in an orthonormal matrix is N; the average squared entry is, therefore, $1/N$; the maximum entry is therefore at least $1/\sqrt{N}$. □

This shows that the basis pair *(Spikes, Sinusoids) yields a most mutually incoherent pair*. For this pair, the sparsity condition leading to ideal atomic decomposition will be most generous. There are other examples of extremal bases, the pair (Spikes, Walsh functions) being an example; but these will seem far less "natural" to those with standard mathematical training.

Underlying Theorem VII.1 is the following uncertainty principle.

Theorem VII.3: Let Φ_1 and Φ_2 be orthonormal bases for 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

$$|T| + |W| \geq \left(1 + M^{-1}\right). \tag{7.1}$$

If we compare this result with the earlier uncertainty principles (Theorems II.1 and VI.3), we see that the general bound (7.1) can be a factor of two away from sharpness in those cases. Its generality can be an advantage in some cases. Thus, mutual incoherence of bases has the following pair of implications.

- No signal can be analyzed in both bases and have simultaneously fewer than about M^{-1} nonzero components from Φ_1 and Φ_2 together.
- A signal which is synthesized from fewer than about M^{-1} components from Φ_1 and M^{-1} components from Φ_2 is decomposed by minimum ℓ^1 atomic decomposition perfectly into those components.

It is curious that M was implicitly identified as heuristically significant by Mallat and Zhang [24] in their initial studies of Matching Pursuit; however, we emphasize that M is relevant here for BP ℓ^1 optimization, rather than Matching Pursuit (greedy single-component extraction).

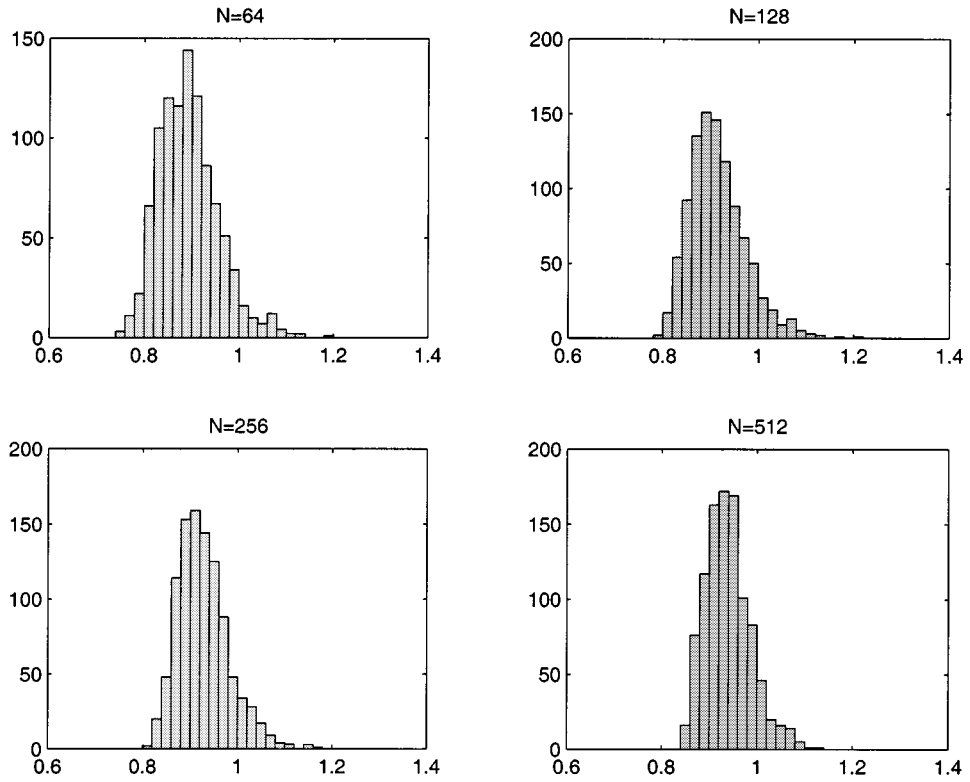

Fig. 1. Empirical distributions of the normalized maximum entry $M(N)/(2\sqrt{\log(N)/N})$ for $N = 64,\ 128,\ 256,\ 512$. Each is based on 1000 simulations.

TABLE I
TABLE OF THE MEDIANS OF THE MAXIMUM AMPLITUDE IN A REAL $N \times N$ PSEUDORANDOM ORTHOGONAL MATRIX

Size N	Median $M(N)$	$2\sqrt{\frac{\log(N)}{N}}$	$\frac{\text{Median } M(N)}{2\sqrt{\log(N)/N}}$
32	0.5684	0.6582	0.8636
64	0.4506	0.5098	0.8837
128	0.3543	0.3894	0.9099
256	0.2706	0.2944	0.9193
512	0.2052	0.2208	0.9296
1024	0.1549	0.1645	0.9413

VIII. Random Orthogonal Bases

To make the point about generality of these results, we now consider random orthogonal bases, their incoherence properties, and some idealized applications.

A. Mutual Incoherence is Generic

Is mutual incoherence special or generic? That is, if one takes a pair of "random orthogonal" bases of R^N, what will be the typical size of M?

The question can be reduced as follows: what is the largest amplitude in a random orthogonal matrix? Here "random" means uniformly distributed on the orthogonal group.

The largest entry in a random real orthogonal matrix is not typically larger than

$$\approx 2\sqrt{\log_e(N)}/\sqrt{N}.$$

We illustrate this in Table I of results based on generation of 100 pseudorandom orthogonal matrices.

Actually, empirical results seem to suggest that the normalized maximum amplitudes $M(N)/(2\sqrt{\log(N)/N})$ converge to a limiting distribution. Fig. 1 gives the empirical distribution out of 1000 simulations.

For a formal result, we have the following.

Theorem VIII.1: Let U denote a random real orthogonal matrix, uniformly distributed on $O(N)$. Let $\epsilon > 0$. Then the exceedance probability

$$\pi_{\epsilon,\,N} = P\left(\max_{ij} |U_{ij}| > 2\sqrt{\log(N)/N}(1+\epsilon)\right)$$

obeys $\pi_{\epsilon,\,N} \to 0$ as $N \to \infty$.

Proof: Any fixed column of a random orthonormal matrix, viewed as a vector in $\boldsymbol{R}^N$, is uniformly distributed on the N-sphere. Each entry $U_{i,\,j}$ can therefore be identified with the projection on the ith coordinate of a randomly chosen point $(U_{i,\,j})_j$ on the N-sphere. This is an exceptionally well-studied quantity; it is the classical example of so-called "concentration of measure phenomena" and "isoperimetry" [21]. It is known that there is very little chance that a random point on the sphere falls far away from the equator; in fact, most distributional properties are similar to those which would hold for a Normally distributed quantity having mean zero and variance $1/N$. Reference [21, Theorem 1.1, p. 15] implies

$$P\left(|U_{ij}| > \zeta/\sqrt{N-2}\right) \le 2\cdot\exp\{-\zeta^2/2\}, \qquad \zeta > 0.$$

From Boole's inequality

$$P\left(\text{any } |U_{ij}| > \zeta/\sqrt{N}\right) \le \sum_{ij} P\left(|U_{ij}| > \zeta/\sqrt{N}\right)$$
$$\le 2\cdot N^2\cdot\exp\left\{-\frac{N-2}{N}\cdot\zeta^2/2\right\}$$

so that taking $\zeta = 2\sqrt{\log(N)}(1+\epsilon)$ we get $\pi_{\epsilon,\,N} \to 0$. □

In short, the $1/\sqrt{N}$ behavior we saw for the incoherence in the (Spikes, Sinusoids) pair is not far from the generic behavior.

For "most" pairs of orthogonal bases of R^N, there is an uncertainty principle threshold and an ideal atomic decomposition threshold, which are both of order $O(\sqrt{N/\log(N)})$.

B. Application: Error-Correcting Encryption

Here is an amusing application of the use of random orthonormal bases in connection with minimum ℓ^1 methods.

A. D. Wyner [33], [34], [29] has advocated a method of encryption for real-valued discrete-time signals S of length N: form a random orthogonal matrix U, multiply the signal vector by the matrix and get the encryption $E = US$. Transmit the encryption to a remote receiver who knows U, and who decrypts via $S = U^T E$. This is an encryption scheme because the observer of E who does not know U sees only that the marginal distribution of the encrypted vector E is uniform on the sphere of radius $\|S\|$ and so there is no "pattern" in E other than the simple pattern of a uniformly distributed vector on the sphere.

The results of this paper show that we may use minimum ℓ^1-norm decomposition in an overcomplete dictionary to extend this encryption scheme so that it is robust against the possibility of gross errors in transmission or recording. Knowing M, the amplitude of the largest entry in matrix U, we encode a vector of $K < M^{-1}/2$ entries by embedding it in a vector S of length N in scattered locations, with the other entries in the vector being zero. We encrypt S according to Wyner's scheme. We transmit E over a channel prone to a small number of gross errors. The receiver performs minimum ℓ^1 atomic decomposition in a combined dictionary consisting of spikes and columns of U.

This variant of the method is robust against gross errors in the transmission and recording of E. Suppose that $\tilde{E}$ agrees with E except in K entries; i.e., $\tilde{E} = E + Z$ where Z has only K nonzero entries. We may view $\tilde{E}$ as a superposition of K terms from the spike dictionary and K terms from the U dictionary. Because $2K < M^{-1}$, we conclude that minimum ℓ^1 atomic decomposition recovers perfectly both the columns of U that correspond to the transmitted data, and the specific locations where $\tilde{E}$ differs from E. In addition, it recovers precisely the entries in the original signal vector S.

Note that the errors can be really large: in principle, they can have an amplitude 1000 or even 10^6 times as large as the amplitude of the transmitted signal, and perfect recovery will still obtain.

Quite generally, then, we can transmit up to $O(\sqrt{N}/\log(N))$ real numbers encrypted in a vector of length N and be immune to up to $O(\sqrt{N}/\log(N))$ gross errors in the transmission and recording of the encrypted data.

C. Application: Separation of Two Uncoordinated Sources

The mutual incoherence of random orthogonal bases has other potential applications. Suppose that an idealized receiver obtains the superposition of two encoded signals

$$R = E_1 + E_2$$

and the goal is to perfectly separate the two signals. For example, R is an idealized antenna and the E_i are received signals from two transmitters which must use the same frequency band. If we are allowed to use this setup with a preprocessing of signals, we can arrange for perfect separation of signals, in principle, even when they are encoded without coordination and are of radically different amplitudes. The idea is that each E_i is a discrete-time signal of length N which is obtained from encoding a message S_i of at most $K < M^{-1}/2$ nonzero entries by applying a random orthogonal transformation U_i to the message vector. Then, with minimum ℓ^1-norm postprocessing at the receiver, we can separate out the two messages perfectly.

This scheme has several key features.

- Each of the two broadcast signals is encrypted and so not accessible to others, including the operator of the other transmitter.
- The transmitters are uncoordinated. The matrices U_i are generated randomly and independently of each other, and each can be kept secret (say) from the owner of the other. Only the receiver operator would need to know both matrices U_i to perform separation.
- The scheme works perfectly, no matter what the relative sizes of the two signals: it works, in principle, at rather enormous differences in transmitter strength.

In comparison, more typical separation schemes would assign each transmitter a subband quasi-disjoint from the other, which requires coordination; also, they rely on linear methods for separation which work poorly when the signal strengths are very different.

IX. Multiscale Bases With Block-Diagonal Structure

While the argumentation so far has mostly been quite general, and could apply to any pair of bases, a special feature of the analysis so far has been that we had M small for large N; $M = O(N^{-1/2})$. If we consider the broader field of applications, this special feature may be absent: we may have M roughly 1. In that case, the above development is rather useless as is.

Nevertheless, we may still obtain interesting insights by extending the approach developed so far. Suppose we have two orthonormal bases Φ_1 and Φ_2, and consider the capacity defined by the optimization problem

$$(K_\gamma) \quad \min \left\|\Phi_1^T x\right\|_1 + \left\|\Phi_2^T x\right\|_1, \qquad \text{subject to } \langle x, \phi_\gamma \rangle = 1$$

In effect, the previous analysis relied on the fact that the value $\mathrm{Val}(K_\gamma)$ did not depend on γ, or at most weakly so.

In some interesting cases, the capacities $\mathrm{Val}(K_\gamma)$ take widely different values, with the largest values being of order 1 independent of N and with many values much smaller than this; in such an event, the preceding analysis by itself tells us almost nothing of any use. An example of such a case is when Φ_1 is a wavelet basis and Φ_2 is a sinusoid basis; at low frequencies, wavelets and sinusoids are not incoherent, and the associated capacity problem (K_γ) has large values; while the value of the capacity problem (K_γ) tends to zero at high frequencies.

Abstracting this situation, we now consider bases with an interesting *block-diagonal structure*. Informally, the γ-indexes can be grouped in blocks in such a way that values within a block of γ-indexes have almost the same value $\mathrm{Val}(K_\gamma)$, and, in addition, the basis functions in a certain group coming from basis

1 span the same space as the basis functions in a corresponding group for basis 2.

Definition IX.1: A pair of orthonormal bases Φ_1, Φ_2 has **joint block diagonal structure** if the following are true.

- There is an orthogonal direct sum decomposition
$$\boldsymbol{R}^N = X_0 \oplus X_1 \oplus \ldots \oplus X_J.$$
- There is a grouping of indexes $\Gamma_{1,j}$ for basis 1 so that
$$\text{SPAN}(\phi_\gamma\colon \gamma \in \Gamma_{1,j}) = X_j$$
and, similarly, a grouping of indexes $\Gamma_{2,j}$ for basis 2 so that
$$\text{SPAN}(\phi_\gamma\colon \gamma \in \Gamma_{2,j}) = X_j.$$

An example of this kind is a combined dictionary (WAVELETS, SINUSOIDS) which will be explained in detail later. We record a simple observation, without proof.

Lemma IX.2: If a pair of bases has joint block-diagonal structure, then the optimization problems (P_0) and (P_1) separate into a direct sum of subproblems, as follows. Let $S^{(j)}$ be the orthoprojection of S on X_j, let $\Phi^{(j)}$ be the subdictionary formed from ϕ_γ with $\gamma \in \Gamma_{1,j} \cup \Gamma_{2,j}$ and define

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

and

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

Then, if a unique solution to each $(P_{0,j})$ exists, a solution to (P_0) is given by the concatenation of all the individual component solutions. Moreover, if a unique solution to each $(P_{1,j})$ exists, a solution to (P_1) is given by the concatenation of all the individual component solutions.

The next observation is immediate.

Lemma IX.3: In the setting of the previous lemma, let

$$M_j = M(\{\phi_\gamma\colon \gamma \in \Gamma_{1,j}\},\ \{\phi_\gamma\colon \gamma \in \Gamma_{2,j}\})$$

be the blockwise mutual incoherence. Then, if S can be represented as a superposition of $N_{1,j}$ terms from $\Gamma_{1,j}$ and $N_{2,j}$ terms from $\Gamma_{2,j}$, and

$$N_{1,j} + N_{2,j} < \tfrac{1}{2} M_j^{-1}$$

the solutions of each $(P_{0,j})$ and each $(P_{1,j})$ are unique and are the same.

For our application, consider a dictionary for discrete-time signals $S(t)$, $t = 0, 1, \ldots, N-1$, made by merging the periodized discrete Meyer orthonormal wavelets basis [20] with an orthonormal basis of certain special orthogonal functions, each one made up of four complex sinusoids of similar frequencies which we will call *real bi-sinusoids*.

The wavelets basis is commonly indexed by $\lambda = (j, k, \epsilon)$ where $j \geq j_0$, $k \in \{0, \ldots, 2^{j+1}\}$, and $\epsilon \in \{0, 1\}$. The basis has, for resolution level $j = j_0$ and gender $\epsilon = 0$, a set of periodized Lemarié scaling functions, and, for resolution levels $j = j_0, j_0+1, \ldots, j_1$, and gender $\epsilon = 1$, the Meyer wavelets; we denote any of these by ψ_λ. Here the effective support of ψ_λ, $\lambda = (j, k, \epsilon)$, is roughly of width $N/2^j$ and so j measures scale.

The real bi-sinusoids e_w are certain special functions, deriving from the construction of the Meyer–Lemarié wavelets. 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. $e_\omega(t) = b_j(w)\cos(2\pi wt/N) - b_j(w')\cos(2\pi w't/N)$, $w < 2^j \cdot 4/3$, $\sigma = 1$;

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

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

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

Here, w' is the "twin" of w, and obeys

$$\begin{aligned} 2^j - w' &= w - 2^j, & w &\leq 2^j \cdot 4/3 \\ 2^{j+1} - w &= w' - 2^{j+1}, & w &> 2^j \cdot 4/3. \end{aligned}$$

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

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

The system e_ω has been constructed so that it is orthonormal and spans the same space W_j as the collection of periodized Meyer wavelets. We call the e_ω real bi-sinusoids because they are made from pairs of real sinusoids.

The key property relating our two bases for W_j can be summarized as follows.

Lemma IX.4: 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.

Proof: By consulting [20] or by adapting arguments from [1], one learns that the algorithm for the discrete periodized Meyer wavelet coefficients at level j of a vector x has five steps. The steps are (for terminology see the cited references)

PMT1. Fourier transform the vector x, yielding $\hat{x}$.

PMT2. Separate $\hat{x}$ into its real and imaginary components.

PMT3. To the frequencies at level j apply *folding projection* to the real and imaginary components of $\hat{x}$ separately, with polarities $(+, -)$ and $(-, +)$, respectively, producing two sequences, $(c_l^j)_l$ and $(d_l^j)_l$.

PMT4. Apply the discrete sine transform DST-III to the c^j sequence and the discrete cosine transform DCT-III to the d^j sequence, yielding sequences $\hat{c}^j$ and $\hat{d}^j$.

PMT5. Combine the results, according to a simple formula

$$\begin{aligned} &\text{for} \quad \lambda=(j, k, 1), && \alpha_\lambda = (-1)^{k+1} \cdot (\hat{c}_k^j + \hat{d}_{2^j-k}^j), \\ & && \quad k=0, 1, \ldots, 2^j-1 \\ &\text{for} \quad \lambda=(j, 2^j+k, 1), && \alpha_\lambda = (-1)^{k+1} \cdot (\hat{c}_k^j - \hat{d}_{2^j-k}^j), \\ & && \quad k=0, 1, \ldots, 2^j-1. \end{aligned}$$

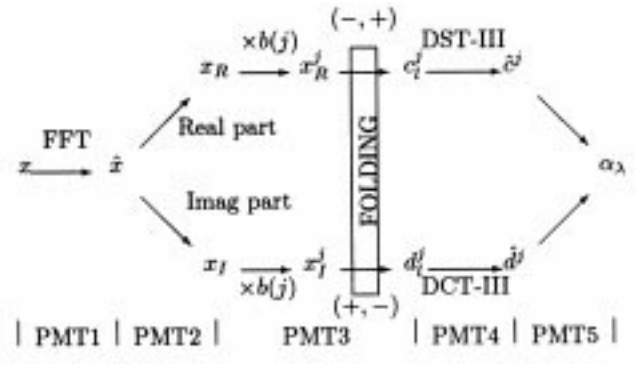

Fig. 2. Flowchart of periodized orthonormal Meyer wavelet transform at scale j.

The key observation is that all these steps are isometries or else isometries up to a scale factor $2^{\pm 1/2}$. It follows that there is an orthonormal basis giving the representers of the output of Step 3. These are exactly the real bi-sinusoids defined earlier

$$c_l^j = \langle x, e_\omega \rangle, \qquad \omega = (l, 1)$$
$$d_l^j = \langle x, e_\omega \rangle, \qquad \omega = (l, 2).$$

In effect, our real bi-sinusoids were obtained by starting from this definition; to obtain formulas RW1–IW2, we started with Kronecker sequences in Ω_j and inverted the transforms in steps PMT3, PMT2, and PMT1.

Now, given this identification, it is clear that the transform U_j mapping real bi-sinusoid coefficients to wavelet coefficients is just the properly scaled composition of steps PMT4 and PMT5, which composition is an isometry. This completes the proof. □

Fig. 2 gives a depiction of the procedure in the above proof.

In short, we have the following structure.

BD1. $\boldsymbol{R^N}$ is partitioned into an orthogonal sum of linear subspaces

$$V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \oplus W_{j_1}.$$

BD2. $\dim(W_j) = 2^{j+1}$.

BD3. Each W_j has two different orthonormal bases: the wavelets $\Psi_{1,j} = (\psi_\lambda : \lambda \in \Lambda_j)$ and the bi-sinusoids $\Psi_{2,j} = (e_\omega : \omega \in \Omega_j)$.

BD4. There is a real orthonormal matrix U_j so that

$$\Psi_{1,j} = U_j \Psi_{2,j}.$$

It follows, upon comparison with Lemma IX.2, that for the combined dictionary

$$\Phi = \Phi_{1,j_0} \cup (\Psi_{1,j_0} \cup \cdots \cup \Psi_{1,j_1}) \cup (\Psi_{2,j_0} \cup \cdots \cup \Psi_{2,j_1})$$

using wavelets at all scales and sinusoids at sufficiently fine scales, the problems (P_0) and (P_1) split into a direct sum of problems $(P_{0,j})$ and $(P_{1,j})$ with $\Phi^{(j)} = \Psi_{1,j} \cup \Psi_{2,j}$, for $j = j_0, j_0+1, \ldots, j_1$, and $S^{(j)}$ the ortho-projection of S onto W_j

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

while

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

with the component $\tilde{S}^{(j_0)}$ of S in V_{j_0} handled by

$$\tilde{S}^{(j_0)} = \sum_{\lambda \in \tilde{\Lambda}_{j_0}} \langle \varphi_\lambda, S \rangle \varphi_\lambda.$$

Lemmas IX.2 and IX.3 draw us to calculate

$$M_j = \sup\{|\langle \psi_{1,\lambda}, \psi_{2,\omega} \rangle|, \lambda \in \Lambda_j, \omega \in \Omega_j\}.$$

This is the mutual incoherence constant M associated with the orthogonal transform between the two bases $\Psi_{1,j}$ and $\Psi_{2,j}$. This will determine the ideal atomic decomposition threshold associated with bases $\Psi_{1,j}$ and $\Psi_{2,j}$.

Lemma IX.5: M_j is exactly the same as the constant for the real Fourier system of cardinality $N = 2^{j-1}$

$$M_j = 2^{-(j-2)/2}.$$

Proof: Let α_j denote the vector of wavelet coefficients at level j and γ_j denote the vector of real bi-sinusoid coefficients at level j stored in order (RW1, RW2, IW1, IW2), then

$$\gamma_j(\omega) = \langle S, e_\omega \rangle, \qquad \omega \in \Omega_j \tag{9.1}$$

and

$$\alpha_j(\lambda) = \langle S, \psi_\lambda \rangle, \qquad \lambda \in \Lambda_j. \tag{9.2}$$

Using column vector notation, we have

$$\alpha_j = \begin{bmatrix} D & \\ & D \end{bmatrix} \begin{bmatrix} \text{(DST-III)} & \text{(DCT-III)} \\ -R\text{(DST-III)} & R\text{(DCT-III)} \end{bmatrix} \gamma_j$$

$$\gamma_j = \frac{1}{2} \begin{bmatrix} \text{(DST-II)} & -\text{(DST-II)}R \\ \text{(DCT-II)} & \text{(DCT-II)}R \end{bmatrix} \begin{bmatrix} D & \\ & D \end{bmatrix} \alpha_j$$

where D is a diagonal matrix

$$D_{kk} = (-1)^{k+1}, \qquad k = 0, 1, 2, \ldots, 2^j - 1$$

(DST-III) and (DST-II) are the matrices of type-III and type-II discrete sine transform; (DCT-III) and (DCT-II) are the matrices of the type-III and type-II discrete cosine transform; and, finally, R is the reversing matrix

$$R = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}.$$

Now the quantity M_j is the amplitude of the largest entry in the matrix representing U_j and obtained by performing the above matrix products. However, by inspection, one sees that M will turn out to be just the largest amplitude in any one of the four submatrices representing the various DC/DS transforms. The closed form for one of these transforms of length N has entries of the form $\sqrt{2/N}$ times a real sinusoid $\cos(\text{argument})$ or $\sin(\text{argument})$ and so we get, by inspection, that the largest entry in such a matrix is not larger than $\sqrt{2/N}$. Taking $N = 2^{j-1}$ we are done. □

And hence we have the following.

Theorem IX.6: Suppose that S is a linear combination of wavelets ψ_λ, $\lambda = (j, k, \varepsilon)$ with $\lambda \in \Lambda$, and of real bi-sinu-

soids e_ω with $\omega \in \Omega$, and the sets of synthesis Λ and Ω obey level-wise the inequality

$$|\Lambda \cap \Gamma_{1,j}| + |\Omega \cap \Gamma_{2,j}| < \frac{1}{2}\left(1 + 2^{(j-2)/2}\right). \qquad (9.3)$$

There is, at most, one way of writing S as such a superposition, and the corresponding sparse vector α is the unique solution of both (P_0) and (P_1).

Some remarks follow.

1) If S obeys the condition (9.3) only at some levels and not others, then at least one can say that decomposition according to (P_0) and (P_1) are identical at all levels where the condition holds.
2) In essence, the subdictionaries are becoming increasingly disjoint as $j \to \infty$, so the sparsity constraint is essentially less restrictive at large j.

No essential role is played in Theorem IX.6 by the finite dimensionality of the overall space $\boldsymbol{R^N}$. Accordingly, we may consider dictionaries with joint block-diagonal structure in the form of infinite direct sums and reach similar conclusions.

In fact, there is a simple dictionary of this form, based on Meyer wavelets on the continuum circle $[0, 2\pi)$ and real bi-sinusoids on the continuum circle. Without going into details, which are exactly parallel to those in the discrete-time case above (see [1]), we get a sequence of vector spaces $\tilde{V}_{j_0}$ and $\tilde{W}_j$, $j \geq j_0$ obeying

$$L^2[0, 2\pi) = \tilde{V}_{j_0} \oplus \tilde{W}_{j_0} \oplus \tilde{W}_{j_0+1} \oplus \cdots$$

and each of these is spanned by basis functions in the corresponding groupings. Continuing in this way, we would reach similar conclusions to Theorem IX.6: under the sparsity conditions

$$|\Lambda \cap \Gamma_{1,j}| + |\Omega \cap \Gamma_{2,j}| < C \cdot 2^{j/2}, \qquad j = j_0, j_0+1, \ldots$$

there is, at most, one way of writing a function in L^2 obeying that condition and the minimum-l^1-norm decomposition finds it.

X. Multiscale Bases With Band Structure

A drawback of Theorem IX.6 and the extension to $L^2[0, 2\pi)$ is that the real bi-sinusoids are not classical sinusoids. At first blush, one thinks to use the fact that each real bi-sinusoid is a sum of two real sinusoids which implies, in an obvious notation

$$N_j(\text{REAL BI-SINUSOIDS}) \leq \sum_{|j-j'|\leq 1} N_{j'}(\text{REAL SINUSOIDS}).$$

It follows that if the object f is a superposition of wavelets and real sinusoids, but we use a dictionary of wavelets and real bi-sinusoids, then under the sparsity condition

$$N_j(\text{WAVELETS}) + N_j(\text{REAL SINUSOIDS}) \leq C2^{j/2}, \qquad j = j_0+1, j_0+2, \ldots$$

the decomposition into wavelets and real bi-sinusoids is unique according to (P_0) and (P_1), involving only the precise wavelets occurring in the expansion of f and the precise real bi-sinusoids appearing in the expansion of sinusoids by real bi-sinusoids. However, it seems to us that a conceptually cleaner result is Theorem I.3 of the Introduction, which assumes that f is made from wavelets and sinusoids and the dictionary is made from wavelets and sinusoids. For a result of that form, we generalize somewhat from block-diagonal structure to band structure.

A further example of ideal atomic decomposition can be given in a multiscale setting where basis Φ_1 is associated with a multiresolution decomposition

$$L^2[0, 2\pi) = V^1_{j_0} \oplus W^1_{j_0} \oplus \cdots \oplus W^1_j \oplus W^1_{j+1} \oplus \cdots$$

and basis Φ_2 is associated with another multiresolution decomposition

$$L^2[0, 2\pi) = V^2_{j_0} \oplus W^2_{j_0} \oplus \cdots \oplus W^2_j \oplus W^2_{j+1} \oplus \cdots.$$

But now $W^1_j \neq W^2_j$. The new condition is the *block-bandedness*

$$W^1_j \cap W^2_{j'} = \emptyset, \qquad |j - j'| > h. \qquad (10.1)$$

Consider the following formal structure.

BB1. $\Phi = \Phi_1 \cup \Phi_2$.

BB2. The index set Γ_1 for the atoms in Φ_1 can be partitioned into subsets $\Gamma_{1,j}$ with $W^1_j = \text{SPAN}\{\varphi_\gamma\colon \gamma \in \Gamma_{1,j}\}$. And similarly for Φ_2, the index set Γ_2 for the atoms in Φ_2 can be partitioned into subsets $\Gamma_{2,j}$ with $W^2_j = \text{SPAN}\{\varphi_\gamma\colon \gamma \in \Gamma_{2,j}\}$.

BB3. For the capacity

$$K(\gamma) = \inf\left\{\left\|\alpha^1\right\|_1 + \left\|\alpha^2\right\|_1\right\}$$

$$\text{subject to } \begin{cases} \langle f, \varphi_\gamma\rangle = 1 \\ \alpha^1 = (\langle f, \varphi_\gamma\rangle\colon \gamma \in \Gamma_1) \\ \alpha^2 = (\langle f, \varphi_\gamma\rangle\colon \gamma \in \Gamma_2) \end{cases}$$

we have the *level-wise capacity*

$$C(j) = \inf K(\gamma), \qquad \text{subject to } \gamma \in \Gamma_{1,j} \cup \Gamma_{2,j}$$

obeying the crucial condition

$$C(j) \to +\infty, \qquad \text{as } j \to +\infty.$$

BB4. We have the block-bandedness (10.1).

Lemma X.1: In the setting *BB1–BB4*, there exists a sequence of critical numbers $N_j \to +\infty$ with the following interpretation. If an L^2 function $f = \sum_{\gamma\in\Gamma} \alpha_\gamma \varphi_\gamma$ is made of a countable number of atoms from Γ with

$$\#\{\gamma \in \Gamma_{1,j} \cup \Gamma_{2,j}\} < N_j \qquad (10.2)$$

then:

a) there is at most one way in which f can be decomposed into a sum of atoms obeying this sparsity condition;
b) the minimum l^1-norm decomposition (P_1) has a unique solution;
c) the solution is the unique decomposition obeying (10.2); and
d) we may take

$$N_j = \frac{1}{4h+2} \min_{|j-j'|\leq h} C(j'). \qquad (10.3)$$

Proof: Let Γ_0 collect the indexes of atoms of f appearing nontrivially in a decomposition $f = \sum_{\gamma\in\Gamma} \alpha_\gamma\varphi_\gamma$. By hypothesis, let Γ_0 have, at most, a finite number of terms α_γ from each $\Gamma_{1,j}$ and $\Gamma_{2,j}$. We are interested in showing that for any object g having coefficients $x^1_\gamma = \langle g, \varphi_\gamma\rangle$, $\gamma \in \Gamma_1$ and x^2_γ in basis 2

$$\sum_{\Gamma_0\cap\Gamma_1} |x^1_\gamma| + \sum_{\Gamma_0\cap\Gamma_2} |x^2_\gamma| < \frac{1}{2}\left(\left\|(x^1_\gamma)\right\|_1 + \left\|(x^2_\gamma)\right\|_1\right). \quad (10.4)$$

It will follow, as before, that the minimum ℓ^1-norm atomic decomposition is precisely $f = \sum_{\Gamma_0} \alpha_\gamma\varphi_\gamma$.

Now for $\gamma \in \Gamma_{1,j}$

$$|x^1_\gamma| \le C(j)^{-1}\cdot\left(\|x^1\|_1 + \|x^2\|_1\right)$$

and, similarly, for $\gamma \in \Gamma_{2,j}$

$$|x^2_\gamma| \le C(j)^{-1}\cdot\left(\|x^1\|_1 + \|x^2\|_1\right).$$

For a vector $x^1 = (x^1_\gamma\colon \gamma \in \Gamma_1)$, let $x^{1,j} = (x^1_\gamma\colon \gamma \in \Gamma_{1,j})$ and similarly for $x^{2,j}$. Let $\pi_{1,j} = \|x^{1,j}\|_1/(\|x^1\|_1 + \|x^2\|_1)$ and similarly for $\pi_{2,j}$. Then from the short-range interaction between scales (10.1)

$$|x^1_\gamma| \le C(j)^{-1}\cdot\left(\sum_{|j'-j|\le h}(\pi_{1,j'}+\pi_{2,j'})\right)\cdot\left(\|x^1\|_1+\|x^2\|_1\right)$$

and, similarly, for x^2_γ. It follows, letting

$$\Gamma_{0,j} = \Gamma_0 \cap (\Gamma_{1,j}\cup\Gamma_{2,j})$$

that

$$\sum_j\sum_{\Gamma_{0,j}}|x_\gamma| \le \sum_j C(j)^{-1}(\#\Gamma_{0,j})\cdot\sum_{|j'-j|\le h}(\pi_{1,j'}+\pi_{2,j'}) \cdot\left(\|x^1\|_1+\|x^2\|_1\right).$$

Now note that

$$\begin{aligned}\sum_j C(j)^{-1}(\#\Gamma_{0,j})&\sum_{|j'-j|\le h}(\pi_{1,j'}+\pi_{2,j'})\\ &\le \sum_{j'}(\pi_{1,j'}+\pi_{2,j'})\left[\sum_{|j'-j|\le h} C(j)^{-1}\#\Gamma_{0,j}\right]\\ &\le \sup_{j'}\sum_{|j'-j|\le h} C(j)^{-1}\#\Gamma_{0,j};\end{aligned}$$

as $\sum_{j'}(\pi_{1,j'}+\pi_{2,j'}) \le 1$. In short, if

$$\sup_{j'}\sum_{|j'-j|\le h} C(j)^{-1}\#\Gamma_{0,j} < \frac{1}{2},$$

the sufficient condition (10.4) will follow. Now if, as in (10.3), $\#\Gamma_{0,j} < \frac{1}{4h+2}C(j)$, then

$$\sup_{j'}\sum_{|j'-j|\le h} C(j)^{-1}\#\Gamma_{0,j} < \sup_{j'}\sum_{|j'-j|\le h} C(j)^{-1}\frac{1}{4h+2}C(j) = \frac{1}{2}.$$

This completes the proof. □

We now consider a dictionary built from an orthobasis of Meyer wavelets combined with an orthobasis of true sinusoids. In this case, the V_{j_0} and W_j are just as in the previous section, but the W'_j are now, simply, the collection of all sines and cosines $\cos(w\theta)$ and $\sin(w\theta)$ with $2^j \le w < 2^{(j+1)}$ (i.e., sinusoids rather than bi-sinusoids). A key point is that since the transformation from real bi-sinusoids to real sinusoids involves only $\Gamma_j, \Gamma_{j'}$ at two adjacent values $|j-j'| \le 1$, it follows that the bandedness condition (10.1) holds with $h = 1$. A second key point is that each $C(j)$ in this case differs from the corresponding $C(j)$ in the real bi-sinusoid case by at most a factor 2.

Combining these observations gives a proof of Theorem I.3 of the Introduction in the case where we interpret SINUSOIDS to mean "real sinusoids."

The proof in the case where we interpret SINUSOIDS to mean complex sinusoids is similar.

XI. Wavelets and Ridgelets

We now turn to Theorem I.4 of the Introduction. This example, combining the Meyer wavelets and orthonormal ridgelets, has a block-banded structure.

We work with functions $f(x_1, x_2)$ in $L^2(\boldsymbol{R}^2)$ and consider two orthonormal sets: for Φ_1 the 2-D Meyer wavelets [1], [25], and for Φ_2 the orthonormal ridgelets [14]. The key properties we use are as follows.

1) The Meyer wavelets have frequency-domain support in a rectangular annulus A^1_j: $\operatorname{supp}\{\psi_{j,k_1,k_2,\varepsilon}(\xi)\}$ satisfies
$$\xi \in \left[-\frac{8}{3}\pi 2^j, \frac{8}{3}\pi 2^j\right]\Big\backslash\left[-\frac{2}{3}\pi 2^j, \frac{2}{3}\pi 2^j\right].$$
2) Orthonormal ridgelets have frequency-domain support in a circular annulus A^2_j: for λ at level j,
$$|\xi| \in \left[\frac{2}{3}\pi 2^j, \frac{8}{3}\pi 2^j\right].$$
3) We use a coarse scale $j_0 > 0$ for the Meyer wavelets, and we use only the part of the ridgelet basis at ridge scales $j > j_0 + 2$.
4) We have $A^1_j \cap A^2_{j'} = \emptyset$ if $\sqrt{2}\frac{8}{3}2^j < \frac{2}{3}2^{j'}$, i.e., $j \le j' - 3$, or $\frac{8}{3}2^{j'} < \frac{2}{3}2^j$, i.e., $j' \le j - 2$.
5) For
$$W^1_j = \text{SPAN}\{\psi_{j,k_1,k_2}\}$$
$$W^2_j = \text{SPAN}\{\varphi_{j,k,i,l}\}$$
$W^1_j \perp W^2_{j'} = \emptyset$, for $|j'-j| > 2$.

In short, we have block bandedness with $h = 2$.

We now calculate the level-wise capacity $C(j)$. We may write

$$K(\gamma) = (1 + 1/\sup\{|\langle\varphi_\gamma, \varphi_{\gamma'}\rangle|\colon \gamma \ne \gamma'\})$$

and

$$C(j) = \inf\{K(\gamma)\colon \gamma \in \Gamma_{1,j}\cup\Gamma_{2,j}\}.$$

The following Lemma shows that $C(j) \le C2^{-j/2}$ and proves Theorem I.4 of the Introduction.

Lemma XI.1: For the wavelet/ridgelet pair, and $\gamma \in \Gamma_{1,j}\cup\Gamma_{2,j}$

$$\sup\{|\langle\psi_{j,k_1,k_2,\varepsilon}, \rho_\lambda\rangle|\colon \lambda\} \le C2^{-j/2}$$

$$\sup\{|\langle\rho_\lambda, \psi_{j,k_1,k_2,\varepsilon}\rangle|\colon (j, k_1, k_2, \varepsilon)\} \le C2^{-j/2}$$

where C is a constant independent of γ.

Proof: We pass to the frequency domain:

$$\begin{aligned}|\langle\psi,\,\rho\rangle| &= \frac{1}{2\pi}\,|\langle\hat{\psi},\,\hat{\rho}\rangle| \\ &\le \frac{1}{2\pi}\,\|\hat{\psi}\|_\infty\|\hat{\rho}\|_1 \\ &= \frac{1}{2\pi}\cdot 2^{-j}\cdot 2^{j/2}\cdot C \\ &= C\cdot 2^{-j/2}.\end{aligned}$$

Here the estimates

$$\begin{aligned}\|\hat{\psi}\|_\infty &\le C2^{-j} \\ \|\hat{\rho}\|_1 &\le C2^{j/2}\end{aligned}$$

follow from known closed-form expressions for $\hat{\psi}$ and $\hat{\rho}$. □

XII. Nonorthogonal Dictionaries

Much of what we have done can be generalized to the case where Φ_1 and Φ_2 are not required to be orthogonal bases. In this case, we measure incoherence via

$$\tilde{M}(\Phi_1,\,\Phi_2) = \max\left[\max_{ij}\left|\Phi_1^{-1}\Phi_2\right|_{ij},\,\max_{ij}\left|\Phi_2^{-1}\Phi_1\right|_{ij}\right]$$

which agrees with the previous measure if Φ_1 and Φ_2 are orthogonal; here, Φ_1^{-1} stands for the matrix inverse to Φ_1 and, similarly, Φ_2^{-1}. We record the essential conclusions.

Theorem XII.1: Let Φ_1 and Φ_2 be bases for $\boldsymbol{R}^N$, and let $\Phi = \Phi_1 \cup \Phi_2$ be the dictionary obtained by merging the two bases. Suppose that S can be represented as a superposition of N_1 atoms from basis 1 and N_2 atoms from basis 2. If

$$N_1 + N_2 \le \frac{1}{2}\tilde{M}(\Phi_1,\,\Phi_2)^{-1}$$

then the solution to (P_1) is unique, the solution to (P_0) is unique, and they are the same.

Proof: With the capacity $(\tilde{K}_\gamma)$ now defined by

$$(\tilde{K}_\gamma)\quad \inf \left\|\Phi_1^{-1}x\right\|_1 + \left\|\Phi_2^{-1}x\right\|_1, \quad \text{subject to } \langle x,\,\varphi_\gamma^*\rangle = 1$$

where φ_γ is a basis function and φ_γ^* is its adjoint. We use the estimate, for γ associated with basis 1

$$\begin{aligned}\langle x,\,\varphi_\gamma^*\rangle &= \langle\Phi_2^{-1}x,\,\Phi_2^T\varphi_\gamma^*\rangle \\ &\le \left\|\Phi_2^{-1}x\right\|_1\left\|\Phi_2^T\varphi_\gamma^*\right\|_\infty.\end{aligned}$$

So, with δ_γ, the Kronecker sequence located at γ

$$\begin{aligned}\left\|\Phi_2^T\varphi_\gamma^*\right\|_\infty &= \left\|\Phi_2^T\Phi_1^{(-1)T}\delta_\gamma\right\|_\infty \\ &\le \max_{\gamma_1\in\Gamma_1,\,\gamma_2\in\Gamma_2}\left|\Phi_2^T\Phi_1^{(-1)T}\right|_{\gamma_1\gamma_2} \\ &\le \tilde{M}.\end{aligned}$$

Hence,

$$\text{VAL}(\tilde{K}_\gamma) \ge 1 + \tilde{M}^{-1}.$$

Arguing as before, this implies that for a subset $\Gamma_0 \subset \Gamma$

$$\begin{aligned}\sum_{\gamma\in\Gamma_0\cap\Gamma_1}\left|\Phi_1^{-1}x\right|_\gamma + \sum_{\gamma\in\Gamma_0\cap\Gamma_2}\left|\Phi_2^{-1}x\right|_\gamma \\ < \left(1+\tilde{M}^{-1}\right)^{-1}|\Gamma_0|\left(\left\|\Phi_1^{-1}x\right\|_1 + \left\|\Phi_2^{-1}x\right\|_1\right).\end{aligned}$$

It follows that if S is generated by atoms in Γ_0 and if

$$|\Gamma_0| \le \frac{1}{2}\tilde{M}^{-1}$$

then the solution to (P_1) is unique; the argument for (P_0) is similar. □

As an application, consider the basis Φ_2 of geometrically decaying sinusoids. Let, for fixed $\rho \in (0,\,1)$, $z = \rho\exp\{2\pi i/N\}$. For $\gamma = (2,\,w)$, let

$$\varphi_\gamma(t) = z^{tw}\cdot\frac{1}{\sqrt{N}}.$$

Then the $\{\varphi_\gamma\colon w = 0,\,1,\,\ldots,\,N-1\}$ are linearly independent but not orthogonal; they would be orthonormal if $\rho = 1$, but we consider only the case $0 < \rho < 1$, which forbids this. With a certain application in mind, we are mainly interested in ρ very close to one, e.g., ρ such that $\rho^N \sim c$, c is substantial (e.g., $1/10$, or $1/4$). We remark that $\varphi_\gamma^*(t) = (\tilde{z})^{tw}\cdot\frac{1}{\sqrt{N}}$ is the dual basis, where $\tilde{z} = \rho^{-1}\exp\{2\pi i/N\}$. Let Φ_1 be the impulse basis, and let $\Phi = \Phi_1 \cup \Phi_2$. Then

$$\tilde{M}(\Phi_1,\,\Phi_2) = \rho^{-N}/\sqrt{N} = c/\sqrt{N}$$

and we conclude that if S is a superposition of spikes and decaying sinusoids, then supposing

$$\#(\text{SPIKES}) + \#(\text{DECAYING SINUSOIDS}) \le \frac{1}{2c}\sqrt{N}$$

the minimum ℓ^1 atomic decomposition will find the common solution of (P_0) and (P_1).

This should be of interest in magnetic resonance spectroscopy [18], where the recorded signal is

$$S(t) = \text{FID}(t) + \varepsilon(t)$$

where the free-induction decay (FID) is a sparse superposition of decaying exponentials with the $\varepsilon(t)$ representing gross errors occurring at those moments of time where the FID exceeds the analog-to-digital converter's upper bound. The above result says that if the FID is accurately modeled as having a few oscillations with common decay rate, then it can be perfectly recovered despite gross recording errors of arbitrary amplitude in unknown locations. This is of particular interest in connection with the water-line problem of magnetic resonance spectroscopy, where the oscillations due to water are so large that they cause the FID to overflow in the first few recorded samples.

XIII. Discussion

A. Continuous-Time Uncertainty Principles

The point of view in this paper concerns the sparsity of representation in two bases.

If two bases are mutually incoherent, then no signal can have a highly sparse representation in both bases simultaneously.

In the case of discrete-time signals, this can be tangibly related to time-frequency concentration because the spikes make up a time-domain orthobasis. In the case of continuous-time signals, this "lack of simultaneous sparsity" principle does not seem to connect directly with classical uncertainty principles. Those principles concern the extent to which a continuous-time

function $f = (f(t): t \in \boldsymbol{R})$ can have small support in both time and frequency domains simultaneously.

By restating the argument used in Section III above, we obtain a continuous-time uncertainty principle. We now briefly explain that principle, and then turn to some useful implications for our studies of earlier sections.

Define the Fourier transform by

$$\hat{f}(\omega) = \int f(t) \exp\{-2\pi i \omega t\}\, dt.$$

With the 2π factor in the exponent this is a unitary transformation. For sets $T \subset \boldsymbol{R}$ and $W \subset \boldsymbol{R}$, define the concentration functional

$$\mu_c(T, W) = \sup \left\{ \frac{\int_T |f(t)|\, dt + \int_W |\hat{f}(\omega)|\, d\omega}{\|f\|_{L^1} + \|\hat{f}\|_{L^1}} : f \in L^1 \cap \mathcal{F}L^1 \right\}.$$

This measures the extent to which an integrable function with integrable Fourier transform can be concentrated to the pair (T, W). We then have, by arguments parallel to Section III.

Theorem XIII.1:

$$\mu_c(T, W) \le |T| + |W|.$$

For example, a function cannot have more than 90% of its combined L^1 norms in (T, W) unless $|T| + |W| > 0.9$.

Proof: Define the capacities

$$(K_{1,t}) \quad \inf \|f\|_{L^1} + \|\hat{f}\|_{L^1}: \qquad f(t) = 1.$$

Evidently, $\mathrm{Val}(K_{1,t})$ is independent of t. Similarly, define

$$(K_{2,\omega}) \quad \inf \|f\|_{L^1} + \|\hat{f}\|_{L^1}: \qquad \hat{f}(\omega) = 1$$

and note also that $\mathrm{Val}(K_{2,\omega})$ is independent of ω. From the completely interchangeable roles of time and frequency, $\mathrm{Val}(K_{2,0}) = \mathrm{Val}(K_{1,0})$. From

$$f(0) = \int \hat{f}(\omega)\, d\omega$$

we have

$$|f(0)| \le \|\hat{f}\|_{L^1}$$

and so $\mathrm{Val}(K_{1,0}) \ge 1$, while setting $f(t) = \exp\{-t^2/\sigma^2\}$ with $\sigma \to 0$ shows that we can have functions f with $f(0) = 1$, $\|f\|_{L^1} \approx 0$, and $\|\hat{f}\|_{L^1} = 1$, hence,

$$\mathrm{Val}(K_{1,0}) = 1.$$

Now if $\|f\|_{L^1} + \|\hat{f}\|_{L^1} = 1$

$$\begin{aligned} \int_T |f(t)|\, dt + \int_W |\hat{f}(\omega)|\, d\omega &\le \int_T \mathrm{Val}(K_{1,t})^{-1}\, dt + \int_W \mathrm{Val}(K_{2,\omega})^{-1}\, d\omega \\ &= \int_T 1\, dt + \int_W 1\, d\omega \\ &= |T| + |W|. \end{aligned}$$

□

This form of uncertainty principle is more symmetric and so, in a sense, more natural than related L^1 uncertainty principles [31] and, of course, it gives the same type of insight.

B. Behavior of μ for Scattered Sets

The connection to the uncertainty principle is useful, above all, for the insights it gives back to the possible behavior of $\mu(T, W)$. It suggests immediately that the sufficient condition $(\mu < 1/2)$ for ideal atomic decomposition holds for many sets T and W where the combined cardinality of T and W far exceeds $\sqrt{N}$, cardinalities as large as $c \cdot N$ being possible, if T and W have the right "disorganization."

In [15], the behavior of a functional similar to the quantity μ_0 of Section VI was studied for a collection of randomly generated, highly scattered sets T, W. Also some basic analysis of simple T, W configurations was made. It was found that if T and W are in some sense "scattered," one could have quite small μ even though T and W were very large sets in total measure. In short, a condition like $|T||W|/N \le 1/2$ was found to be in no way *necessary* for low concentration, unless T and W are very carefully arranged in a "picket-fence" form. In [16], the behavior of a functional similar to μ_0 was analyzed in the case where W is an interval. It was found that T could have very large measure, even proportional to N, and still one could have $\mu_0 \ll 1/2$, provided in each interval of a certain length, there was only a small number of points from T; here, the length of the interval was reciprocal to the size of the frequency band W.

Both of these strands of investigation indicate clearly that the $\sqrt{N}$ threshold and the mutual incoherence property should be viewed simply as worst case measures. Typically, we can relax our quantitative sparsity constraint significantly, and, as long as T and/or W are sufficiently scattered, we will still have favorable concentration ratios.

To investigate this idea, we performed a computational experiment in the (Spikes, Sinusoids) dictionary. As in Section III, we note a simple necessary condition for a sequence α_γ to be a unique solution of the ℓ^1 problem. Suppose the sequence is supported on a set $T \cup W$ with sign sequence $\epsilon_\gamma = \mathrm{sign}(\alpha_\gamma)$. In order for it to be a set of uniqueness for the ℓ^1, it is sufficient that, for all $\delta \in \mathcal{N}$

$$\sum_{T \cup W} \epsilon_\gamma \delta_\gamma < \frac{1}{2} \|\delta\|_1.$$

In our experiment, we generated 1000 sets $T \cup W$ with various N, N_t, and N_w, and calculated by linear programming

$$\tilde{\mu}(T, W; \epsilon) = \sup \sum_\gamma \epsilon_\gamma \delta_\gamma, \quad \text{subject to} \quad \|\delta\|_1 \le 1,\ \delta \in \mathcal{N}.$$

As an example, we computed realizations of $\tilde{\mu}$ for $N = 32$ and

$$N_t = N_w \in \{3, 6, 9, 12, 15, 18, 21, 24\}.$$

Fig. 3 presents a histogram of our results, illustrating that, within a given set of parameters N, N_t, N_w, the obtained values of $\tilde{\mu}$ exhibit a roughly Normal distribution, with increasing values of

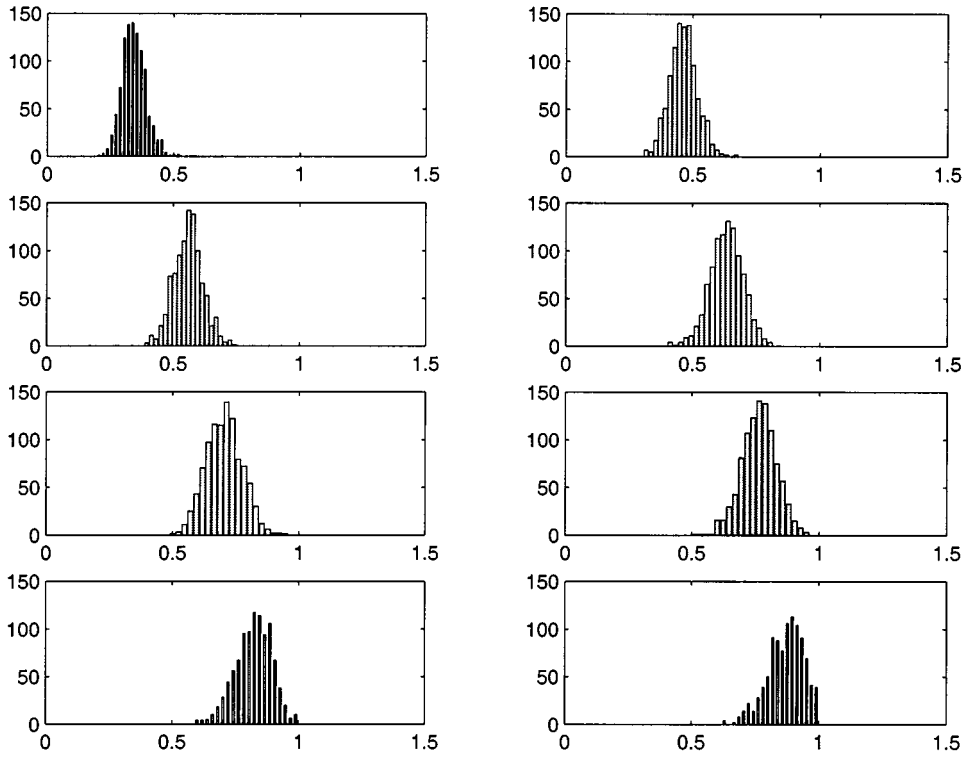

Fig. 3. Histograms of $\tilde{\mu}(T, W; \epsilon)$. $N = 32$. Ordered from left to right and top to bottom, $N_t = N_w = 3, 6, 9, 12, 15, 18, 21, 24$, respectively.

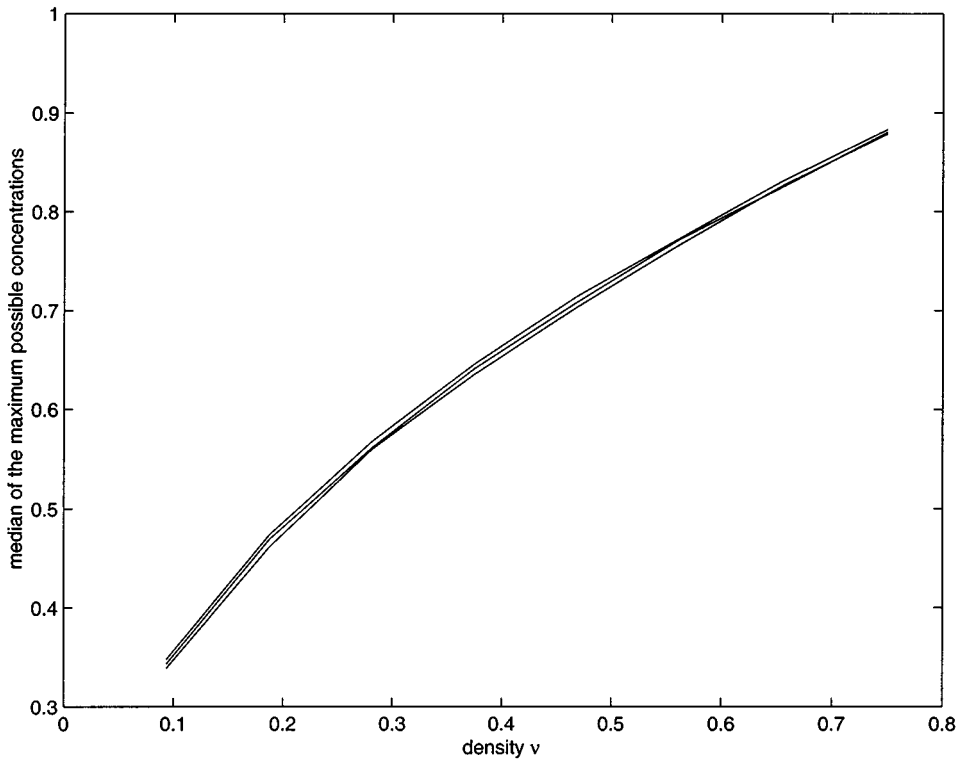

Fig. 4. Plot of $\mathrm{median}\,(\tilde{\mu})$ versus density ν for $N = 32, 64, 128$. The curve associated with $N = 32$ is the lowest and the curve associated with $N = 128$ is the highest.

N_t and N_w leading to increasing $\tilde{\mu}$, as they must. It is clear that a simple numerical summary of the distribution, such as the median of each histogram, will adequately describe the distribution of $\tilde{\mu}$.

Fig. 4 presents a display of the median values of $\tilde{\mu}$ in simulations at $N = 32, 64$ and 128, plotted as three curves, with $\mathrm{median}(\tilde{\mu})$ displayed versus the *density* $\nu = (N_t + N_w)/(2N)$. We make the following observations.

- The curves are very similar at different N, so that the description of $\tilde{\mu}$ as dependent on the density ν seems reasonable.
- The curves are almost linear, roughly obeying the equation

$$\text{median}(\tilde{\mu}) \approx 0.32 + 0.79\nu.$$

- The curves cross the critical threshold concentration $= 1/2$ near $\nu = 0.2$.

These results suggest that for a large collection of triplets (T, W, ϵ) one has, at the same time, $|T| + |W| \sim N/5$ and $\tilde{\mu} < 0.5$; in such cases, the associated (P_1) has a unique solution and the method of minimum ℓ^1-norm atomic decomposition will give a unique solution. This suggests that the results proved in this paper under restrictive sparsity assumptions may point the way to a phenomenon valid under far less restrictive sparsity assumptions.

ACKNOWLEDGMENT

The authors wish to thank S. Mallat for discussions on Matching Pursuit; S. S. Chen and M. Saunders for discussions on Basis Pursuit; and S. Boyd for teaching us about the use of convex optimization to solve relaxed versions of combinatorial optimizations. The referees also provided numerous helpful references and comments.

REFERENCES

[1] P. Auscher, G. Weiss, and M. V. Wickerhauser, "Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets," in *Wavelet*. Boston, MA: Academic Press, 1992, vol. 2, Wavelets Anal. Appl., pp. 237–256.

[2] A. A. Beex and F. G. Safar, "Multiple basis signal representation, coding, and construction," in *IEEE Proc. Int. Symp. Information Threory*, San Diego, CA, Jan. 1990, p. 76.

[3] A. P. Berg and W. B. Mikhael, "A survey of mixed transform techniques for speech and image coding," in *Proc. 1999 IEEE Int. Symp. Circuits and Systems*, vol. 4, 1999, pp. 106–109.

[4] E. J. Candès and D. L. Donoho, "Curvelets," Tech. Rep., Dept. Statistics, Stanford Univ., Stanford, CA, 2001.

[5] ——, "Ridgelets: The key to high-dimensional intermittency?," Phil. Trans. Roy. Soc. London A., 1999, to be published.

[6] S. S. Chen, "Basis pursuit," Ph.D. dissertation, Stanford Univ., Stanford, CA, Nov. 1995.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[8] R. Coifman, Y. Meyer, and V. Wickerhauser, "Adapted wave form analysis, wavelet-packets and applications," in *ICIAM 91. Proc. 2nd Int. Conf. Industrial and Applied Mathematics*. Philadelphia, PA: SIAM, 1992, pp. 41–50.

[9] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[10] I. Daubechies, S. Jaffard, and J.-L. Journé, "A simple Wilson orthonormal basis with exponential decay," *SIAM J. Math. Anal.*, vol. 22, no. 2, pp. 554–573, 1991.

[11] V. E. DeBrunner, L.-X. Chen, and H. J. Li, "Lapped multiple basis algorithms for still image compression without blocking effect," *IEEE Trans. Image Processing*, vol. 6, pp. 1316–1322, 1997.

[12] V. E. DeBrunner, W. Lou, and J. Thripuraneni, "Multiple transform algorithms for time-varying signal representation," *IEEE Trans. Circuits Syst.*, vol. 44, pp. 663–667, 1997.

[13] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 5, no. 2–3, pp. 173–187, 1996.

[14] D. L. Donoho, "Orthonormal ridgelets and linear singularities," *SIAM J. Math Anal.*, vol. 31, no. 5, pp. 1062–1099.

[15] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[16] D. L. Donoho and B. F. Logan, "Signal recovery and the large sieve," *SIAM J. Appl. Math.*, vol. 52, no. 2, pp. 577–591, 1992.

[17] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," *J. Fourier Anal. Appl.*, vol. 3, no. 3, pp. 207–238, 1997.

[18] J. C. Hoch and A. S. Stern, *NMR Data Processing*. New York: Wiley-Liss, 1996.

[19] X. Huo, "Sparse image representation via combined transforms," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1999.

[20] E. D. Kolaczyk, "Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data," Ph.D. dissertation, Stanford Univ., Stanford, CA, Aug. 1994.

[21] M. Ledoux and M. Talagrand, *Probability in Banach spaces: Isoperimetry and Processes*. New York: Springer-Verlag, 1991.

[22] B. F. Logan, "Properties of high-pass signals," Ph.D. dissertation, Dept. Elec. Eng., Columbia Univ., New York, 1965.

[23] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.

[24] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[25] Y. Meyer, *Wavelets Algorithms & Applications*. Philadelphia, PA: SIAM, 1993. Translated and revised by R. Ryan.

[26] W. B. Mikhael and A. S. Spanias, "Accurate representation of time-varying signals using mixed transforms with applications to speech," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 329–331, 1989.

[27] T. Przebinda, V. DeBrunner, and M. Ozaydin, "Using a new uncertainty measure to determine optimal bases for signal representations," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 1999, pp. 1365–1368.

[28] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. San Diego, CA: Academic, 1990.

[29] N. J. A. Sloane, "Encrypting by random rotations," in *Cryptography (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 1983, vol. 149, , pp. 71–128.

[30] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. Cambridge, MA: Wellesley-Cambridge, 1994.

[31] D. N. Williams, "New mathematical proof of the uncertainty relation," *Amer. J. Phys.*, vol. 47, no. 7, pp. 606–607, 1979.

[32] M. H. Wright, "The interior-point revolution in constrained optimization," in *High-Performance Algorithms and Software in Nonlinear Optimization*, E. R. DeLoach, A. Murli, P. M. Pardalos, and G. Toraldo, Eds. Norwell, MA: Kluwer, 1998, pp. 359–381. [Online]. Available: http://cm.bell-labs.com/cm/cs/doc/98/4-09.ps.gz.

[33] A. D. Wyner, "An analog scrambling scheme which does not expand bandwidth, Part I: Discrete time," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 261–274, May 1979.

[34] ——, "An analog scrambling scheme which does not expand bandwidth, Part II: Continuous time," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 415–425, July 1979.