Sparsity and Incoherence in Compressive Sampling

Emmanuel Candès[†] and Justin Romberg[♯]

† Applied and Computational Mathematics, Caltech, Pasadena, CA 91125

♯ Electrical and Computer Engineering, Georgia Tech, Atlanta, GA 90332

November 2006

Abstract

We consider the problem of reconstructing a sparse signal $x^0 \in \mathbb{R}^n$ from a limited number of linear measurements. Given m randomly selected samples of Ux^0, where U is an orthonormal matrix, we show that ℓ_1 minimization recovers x^0 exactly when the number of measurements exceeds

$$m \geq \text{Const} \cdot \mu^2(U) \cdot S \cdot \log n,$$

where S is the number of nonzero components in x^0, and μ is the largest entry in U properly normalized: $\mu(U) = \sqrt{n} \cdot \max_{k,j} |U_{k,j}|$. The smaller μ, the fewer samples needed.

The result holds for "most" sparse signals x^0 supported on a fixed (but arbitrary) set T. Given T, if the sign of x^0 for each nonzero entry on T and the observed values of Ux^0 are drawn at random, the signal is recovered with overwhelming probability. Moreover, there is a sense in which this is nearly optimal since any method succeeding with the same probability would require just about this many samples.

1 Introduction

1.1 Sparse recovery from partial measurements

This paper addresses the problem of signal acquisition in a broad setting. We are interested in "sampling" a vector $x^0 \in \mathbb{R}^n$. Instead of observing x^0 directly, we sample a small number m of

transform coefficients of x^0. For an orthogonal matrix[1] U with

$$U^*U = nI, \tag{1.1}$$

these transform coefficients are given by $y^0 = Ux^0$. Of course, if all n of the coefficients y^0 are observed, recovering x^0 is trivial: we simply apply $\frac{1}{n}U^*$ to the vector of observations y^0. Instead, we are concerned with the highly underdetermined case in which only a small fraction of the components of y^0 are actually sampled or observed. Given a subset $\Omega \subset \{1, \ldots, n\}$ of size $|\Omega| = m$, the challenge is to infer the "long" n-dimensional vector x^0 from the "short" m-dimensional vector of observations $y = U_\Omega x^0$, where U_Ω is the $m \times n$ matrix consisting of the rows of U indexed by Ω. In plain English, we wish to solve a system of linear equations in which there are fewer equations than unknowns.

A special instance of this problem was investigated in a recent paper [4], where U is taken as the usual discrete Fourier transform. The main result of this work is that if x^0 is *S-sparse* (at most S of the n components of x^0 are nonzero), then it can be recovered perfectly from on the order of $S \log n$ Fourier-domain samples. The recovery algorithm is concrete and tractable: given the discrete Fourier coefficients

$$y_k = \sum_{t=1}^{n} x^0(t) e^{-i2\pi(t-1)k/n}, \quad k \in \Omega, \tag{1.2}$$

or $y = F_\Omega x^0$ for short, we solve the convex optimization program

$$\min_x \ \|x\|_{\ell_1} \quad \text{subject to} \quad F_\Omega x = y.$$

For a fixed x^0, the recovery is exact for the overwhelming majority of sample sets Ω of size

$$|\Omega| \geq C \cdot S \cdot \log n, \tag{1.3}$$

where C is a known (small) constant.

Since [4], a theory of "compressed sensing" has developed around several papers [6, 7, 9] demonstrating the effectiveness of ℓ_1 minimization for recovering sparse signals from a limited number of measurements. To date, most of this effort has been focused on systems which take completely unstructured, noise-like measurements, i.e. the observation vector y is created from a series of inner products against random test vectors $\{\phi_k\}$:

$$y_k = \langle \phi_k, x^0 \rangle, \quad k = 1, \ldots, m. \tag{1.4}$$

The collection $\{\phi_k\}$ is sometimes referred to as a *measurement ensemble*; we can write (1.4) compactly as $y = \Phi x^0$, where the rows of Φ are the ϕ_k. Published results take ϕ_k to be a realization of Gaussian white noise, or a sequence of Bernoulli random variables taking values ± 1 with equal probability. This work has shown that taking random measurements is in some sense an optimal strategy for acquiring sparse signals; it requires a near-minimal number of measurements [1,6,7,9,10]

---

[1]On a first reading, our choice of normalization of U may seem a bit strange. The advantages of taking the row vectors of U to have Euclidean norm $\sqrt{n}$ are that 1) the notation in the sequel will be cleaner, and 2) it will be easier to see how this result generalizes the special case of incomplete sampling in the Fourier domain presented in [4].

— m measurements can recover signals with sparsity $S \lesssim m/\log(n/m)$, and all of the constants appearing in the analysis are small [13]. Similar bounds have also appeared using greedy [28] and complexity-based [17] recovery algorithms in place of ℓ_1 minimization.

Although theoretically powerful, the practical relevance of results for completely random measurements is limited in two ways. The first is that we are not always at liberty to choose the types of measurements we use to acquire a signal. For example, in magnetic resonance imaging (MRI), subtle physical properties of nuclei are exploited to collect samples in the Fourier domain of a two- or three-dimensional object of interest. While we have control over which Fourier coefficients are sampled, the measurements are inherently frequency based. A similar statement can be made about tomographic imaging; the machinery in place measures Radon slices, and these are what we must use to reconstruct an image.

The second drawback to completely unstructured measurement systems is computational. Random (i.e. unstructured) measurement ensembles are unwieldy numerically; for large values of m and n, simply storing or applying Φ (tasks which are necessary to solve the ℓ_1 minimization program) are nearly impossible. If, for example, we want to reconstruct a megapixel image ($n = 1,000,000$) from $m = 25,000$ measurements (see the numerical experiment in Section 2), we would need more than 3 gigabytes of memory just to store the measurement matrix, and on the order of gigaflops to apply it. The goal from this point of view, then, is to have similar recovery bounds for measurement matrices Φ which can be applied quickly (in $O(n)$ or $O(n \log n)$ time) and implicitly (allowing us to use a "matrix free" recovery algorithm).

Our main theorem, stated precisely in Section 1.2 and proven in Section 3, states that bounds analogous to (1.3) hold for sampling with general orthogonal systems. We will show that for a fixed signal support T of size $|T| = S$, the program

$$\min_x \ \|x\|_{\ell_1} \quad \text{subject to} \quad U_\Omega x = U_\Omega x^0 \tag{1.5}$$

recovers the overwhelming majority of x^0 supported on T and observation subsets Ω of size

$$|\Omega| \geq C \cdot \mu^2(U) \cdot S \cdot \log n, \tag{1.6}$$

where $\mu(U)$ is simply the largest magnitude among the entries in U:

$$\mu(U) = \max_{k,j} \ |U_{k,j}|. \tag{1.7}$$

It is important to understand the relevance of the parameter $\mu(U)$ in (1.6). $\mu(U)$ can be interpreted as a rough measure of how concentrated the rows of U are. Since each row (or column) of U necessarily has an ℓ_2-norm equal to $\sqrt{n}$, μ will take a value between 1 and $\sqrt{n}$. When the rows of U are perfectly flat — $|U_{k,j}| = 1$ for each k, j, as in the case when U is the discrete Fourier transform, we will have $\mu(U) = 1$, and (1.6) is essentially as good as (1.3). If a row of U is maximally concentrated — all the row entries but one vanish — then $\mu^2(U) = n$, and (1.6) offers us no guarantees for recovery from a limited number of samples. This result is very intuitive. Suppose indeed that $U_{k_0,j_0} = \sqrt{n}$ and x^0 is 1-sparse with a nonzero entry in the j_0th location. To reconstruct x^0, we need to observe the k_0th entry of Ux^0 as otherwise, the data vector y will vanish. In other words, to reconstruct x^0 with probability greater than $1 - 1/n$, we will need to see all the

3

components of Ux^0, which is just about the content of (1.6). This shows informally that (1.6) is fairly tight on both ends of the range of the parameter μ.

For a particular application, U can be decomposed as a product of a *sparsity basis* Ψ, and an orthogonal measurement system Φ. Suppose for instance that we wish to recover a signal $f \in \mathbb{R}^n$ from m measurements of the form $y = \Phi f$. The signal may not be sparse in the time domain but its expansion in the basis Ψ may be

$$f(t) = \sum_{j=1}^{n} x_j^0 \psi_j(t), \quad f = \Psi x$$

(the columns of Ψ are the discrete waveforms ψ_j). Our program searches for the coefficient sequence in the Ψ-domain with minimum ℓ_1 norm that explains the samples in the measurement domain Φ. In short, it solves (1.6) with

$$U = \Phi\Psi, \quad \Psi^*\Psi = I, \quad \Phi^*\Phi = nI.$$

The result (1.6) then tells us how the relationship between the sensing modality (Φ) and signal model (Ψ) affects the number of measurements required to reconstruct a sparse signal. The parameter μ can be rewritten as

$$\mu(\Phi\Psi) = \max_{k,j} |\langle \phi_k, \psi_j \rangle|,$$

and serves as a rough characterization of the degree of similarity between the sparsity and measurement systems. For μ to be close to its minimum value of 1, each of the measurement vectors (rows of Φ) must be "spread out" in the Ψ domain. To emphasize this relationship, $\mu(U)$ is often referred to as the *mutual coherence* [11, 12]. The bound (1.6) tells us that an S-sparse signal can be reconstructed from $\sim S \log n$ samples in any domain in which the test vectors are "flat", i.e. the coherence parameter is $O(1)$.

1.2   Main result

The ability of the ℓ_1-minimization program (1.5) to recover a given signal x^0 depends only on 1) the support set T of x^0, and 2) the sign sequence z_0 of x^0 on T.[2] For a fixed support T, our main theorem shows that perfect recovery is achieved for the overwhelming majority of the combinations of sign sequences on T, and sample locations (in the U domain) of size m obeying (1.6).

The language "overwhelming majority" is made precise by introducing a probability model on the set Ω and the sign sequence z. The model is simple: select Ω uniformly at random from the set of all subsets of the given size m; choose each $z(t)$, $t \in T$ to be ± 1 with probability $1/2$. Our main result is:

Theorem 1.1 *Let U be an $n \times n$ orthogonal matrix ($U^*U = nI$) with $|U_{k,j}| \leq \mu(U)$. Fix a subset T of the signal domain. Choose a subset Ω of the measurement domain of size $|\Omega| = m$, and a sign sequence z on T uniformly at random. Suppose that*

$$m \geq C_0 \cdot |T| \cdot \mu^2(U) \cdot \log(n/\delta) \tag{1.8}$$

---

[2]In other words, the recoverability of x^0 is determined by the *facet* of the ℓ_1 ball of radius $\|x^0\|_{\ell_1}$ on which x^0 resides.

and also $m \geq C_0' \cdot \log^2(n/\delta)$ for some fixed numerical constants C_0 and C_0'. Then with probability exceeding $1-\delta$, every signal x^0 supported on T with signs matching z can be recovered from $y = U_\Omega x^0$ by solving (1.5).

The hinge of Theorem 1.1 is a new *weak uncertainty principle* for general orthobases. Given T and Ω as above, it is impossible to find a signal which is concentrated on T and on Ω in the U domain. In the example above where $U = \Phi\Psi$, this says that one cannot be concentrated on small sets in the Ψ and Φ domains simultaneously. As noted in previous publications [3,4], this is a statement about the eigenvalues of minors of the matrix U. Let U_T be the $n \times |T|$ matrix corresponding to the columns of U indexed by T, and let $U_{\Omega T}$ be the $m \times |T|$ matrix corresponding to the rows of U_T indexed by Ω. In Section 3, we will prove the following:

Theorem 1.2 *Let U, T, and Ω be as in Theorem 1.1. Suppose that the number of measurements m obeys*

$$m \geq |T| \cdot \mu^2(U) \cdot \max(C_1 \log |T|, \ C_2 \log(3/\delta)), \tag{1.9}$$

for some positive constants C_1, C_2. Then

$$\mathbf{P}\left(\|\frac{1}{m}U_{\Omega T}^* U_{\Omega T} - I\| \geq 1/2 \right) \leq \delta, \tag{1.10}$$

where $\| \cdot \|$ is the standard operator ℓ_2 norm—here, the largest eigenvalue (in absolute value).

For small values of δ, the eigenvalues of $U_{\Omega T}^* U_{\Omega T}$ are all close to m with high probability. To see that this is an uncertainty principle, let $x \in \mathbb{R}^n$ be a sequence supported on T, and suppose that $\|m^{-1}U_{\Omega T}^* U_{\Omega T} - I\| \leq 1/2$. It follows that

$$\frac{m}{2}\|x\|_{\ell_2}^2 \leq \|U_\Omega x\|_{\ell_2}^2 \leq \frac{3m}{2}\|x\|_{\ell_2}^2, \tag{1.11}$$

which asserts that only a small portion of the energy of x will be concentrated on the set Ω in the U-domain (the total energy obeys $\|Ux\|_{\ell_2}^2 = n\|x\|_{\ell_2}^2$). Moreover, this portion is essentially proportional to the size of Ω.

1.3   Contributions and relationship to prior work

The relationship of the mutual incoherence parameter μ to the performance of ℓ_1 minimization programs with equality constraints first appeared in the context of *Basis Pursuit* for sparse approximation, see [12] and also [11, 14, 16].

As mentioned in the previous section, [4] demonstrated the effectiveness of ℓ_1 recovery from Fourier-domain samples in slightly more general situations than in Theorem 1.1 (randomization of the signs on T is not required). Obviously, the results presented in this paper considerably extend this Fourier sampling theorem.

We also note that since [4], several papers have appeared on using ℓ_1 minimization to recover sparse signals from a limited number of measurements [5, 7, 9]. In particular [7] and [25] provide bounds for reconstruction from a random subset of measurements selected from an orthogonal basis; these

papers ask that *all* sparse signals to be simultaneously recoverable from the same set of samples (which is stronger than our goal here), and their bounds have log factors of $(\log n)^6$ and $(\log n)^5$ respectively. These results are based on *uniform* uncertainty principles, which require (1.10) to hold for all sets T of a certain size simultaneously once Ω is chosen. Whether or not this log power can be reduced in this context remains an open question.

A contribution of this paper is to show that if one is only interested in the recovery of nearly all signals on a fixed set T, these extra log factors can indeed be removed. We show that to guarantee exact recovery, we only require $U_{\Omega T}$ to be well behaved for this fixed T as opposed to all T's of the same size, which is a significantly weaker requirement. By examining the singular values of $U_{\Omega T}$, one can check whether or not (1.11) holds.

Our method of proof, as the reader will see in Section 3, relies on a variation of the powerful results presented in [24] about the expected spectral norm of certain random matrices. We also introduce a novel large-deviation inequality, similar in spirit to those reviewed in [19,20] but carefully tailored for our purposes, to turn this statement about expectation into one about high probability.

Finally, we would like to contrast this work with [29], which also draws on the results from [24]. First, there is a difference in how the problem is framed. In [29], the $m \times n$ measurement system is fixed, and bounds for perfect recovery are derived when the support and sign sequence are chosen at random, i.e. a fixed measurement system works for most signal supports of a certain size. In this paper, we fix an arbitrary signal support, and show that we will be able to recover from most sets of measurements of a certain size in a fixed domain. Second, although slightly more general class of measurement systems is considered in [29], the final bounds for sparse recovery in the context of (1.5) do not fundamentally improve on the uniform bounds cited above; [29] draws weaker conclusions since the results are not shown to be universal in the sense that all sparse signals are recovered as in [7] and [25].

2 Applications

In the 1990s, image compression algorithms were revolutionized by the introduction of the wavelet transform. The reasons for this can be summarized with two major points: the wavelet transform is a much sparser representation for photograph-like images than traditional Fourier-based representations, and it can be applied and inverted in $O(n)$ computations.

To exploit this wavelet-domain sparsity in acquisition, we must have a measurement system which is incoherent with the wavelet representation (so that μ in (1.6) is small) and that can be applied quickly and implicitly (so that large-scale recovery is computationally feasible). In this section, we present numerical experiments for two such measurement strategies.

2.1 Fourier sampling of sparse wavelet subbands

Our first measurement strategy takes advantage of the fact that at fine scales, wavelets are very much spread out in frequency. We will illustrate this in 1D; the ideas are readily applied to 2D image.
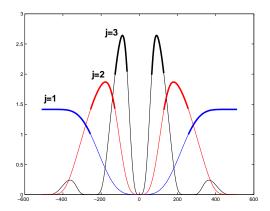
Figure 1: *Wavelets in the frequency domain. The curves shown above are the magnitude of the discrete Fourier transform (2.1) of Daubechies-8 wavelets for $n = 1024$ and $j = 1, 2, 3$. The magnitude of $\hat{\psi}_{j,\cdot}$ over the subband (2.3) is shown in bold.*

Labeling the scales of the wavelet transform by $j = 1, 2, \ldots, J$, where $j = 1$ is the finest scale and $j = J$ the coarsest, the wavelets[3] $\psi_{j,k}$ at scale j are almost flat in the Fourier domain over a band of size $n_j = n2^{-j}$. The magnitude of the Fourier transform

$$\hat{\psi}_{j,k}(\omega) = \sum_{t=1}^{n} \psi(t) e^{-i2\pi(t-1)\omega/n}, \quad \omega = -n/2 + 1, \ldots, n/2, \tag{2.1}$$

is the same for each wavelet at scale j, since

$$\hat{\psi}_{j,k}(\omega) = e^{-i2\pi(k-1)\omega/n_j} \hat{\psi}_{j,1}(\omega). \tag{2.2}$$

These spectrum magnitudes are shown for the Daubechies-8 wavelet in Figure 1. We see that over frequencies in the jth subband

$$\omega \in \mathcal{B}_j := \{n_j/2 + 1, \ldots, n_j\} \cup \{-n_j + 1, \ldots, -n_j/2\}, \tag{2.3}$$

we have

$$\frac{\max_{\omega \in \mathcal{B}_j} |\hat{\psi}_{j,k}(\omega)|}{\min_{\omega \in \mathcal{B}_j} |\hat{\psi}_{j,k}(\omega)|} < \text{Const} \approx \sqrt{2}.$$

Suppose now that a signal x^0 is a superposition of S wavelets at scale j, that is, we can write

$$x^0 = \Psi_j w^0$$

where $w^0 \in \mathbb{R}^{n_j}$ is S-sparse, and Ψ_j is the $n \times n_j$ matrix whose columns are the $\psi_{j,k}(t)$ for $k = 1, \ldots, n_j$. We will measure x^0 by selecting Fourier coefficients from the band $\mathcal{B}_j$ at random. To see how this scheme fits into the domain of the results in the introduction, let ω index the subband $\mathcal{B}_j$, let F_j be the $n_j \times n$ matrix whose rows are the Fourier vectors for frequencies in $\mathcal{B}_j$, let D_j be a diagonal matrix with

$$(D_j)_{\omega,\omega} = \hat{\psi}_{j,1}(\omega), \quad \omega \in \mathcal{B}_j,$$

---

[3]Wavelets are naturally parameterized by a scale j and a shift k with $k = 1, 2, \ldots, n2^{-j}$ — see [21]. The wavelets at a set scale are just circular shifts of one another: $\psi_{j,k}(t) = \psi_{j,1}(t - 2^j k)$, where the substraction is modulo n.

Table 1: *Number of measurements required to reconstruct a sparse subband. Here, $n = 1024$, S is the sparsity of the subband, and $M(S, j)$ is the smallest number of measurements so that the S-sparse subband at wavelet level j was recovered perfectly in 1000/1000 trials.*

$j = 1$		$j = 2$		$j = 3$	
S	$M(S, j)$	S	$M(S, j)$	S	$M(S, j)$
50	100	25	56	15	35
25	68	15	40	8	24
15	49	8	27	-	-

and consider the $n_j \times n_j$ system

$$U = D_j^{-1} F_j \Psi_j.$$

The columns of $F_j \Psi_j$ are just the Fourier transforms of the wavelets given in (2.2),

$$(F_j \Psi_j)_{\omega, k} = e^{-i2\pi(k-1)\omega/n_j} \hat{\psi}_{j,1}(\omega) \quad \Rightarrow \quad U_{\omega, k} = e^{-i2\pi(k-1)\omega/n_j},$$

and so U is just a $n_j \times n_j$ Fourier system. In fact, one can easily check that $U^* U = U^* U = n_j I$.

We choose a set of Fouier coefficients Ω of size m in the band $\mathcal{B}_j$, and measure

$$y = F_\Omega x_0 = F_\Omega \Psi_j w^0,$$

which can easily be turned into a set of samples in the U domain $y' = U_\Omega w^0$ just by re-weighting y. Since the mutual incoherence of $D^{-1} F_j \Psi_j$ is $\mu = 1$, we can recover w^0 from $\sim S \log n$ samples.

Table 1 summarizes the results of the following experiment: Fix the scale j, sparsity S, and a number of measurements m. Perform a trial for (S, j, m) by first generating a signal support T of size S, a sign sequence on that support, and a measurement set Ω_j of size m uniformly at random, and then measuring $y = F_{\Omega_j} \Psi_j x^0$ (x^0 is just the sign sequence on T and zero elsewhere), solving (1.5), and declaring success if the solution matches x^0. A thousand trials were performed for each (S, j, m). The value $M(S, j)$ recorded in the table is the smallest value of m such that the recovery was successful in all 1000 trials. As with the partial Fourier ensemble (see the numerical results in [4]), we can recover from $m \approx 2S$ to $3S$ measurements.

To use the above results in an imaging system, we would first separate the signal/image into wavelet subband, measure Fourier coefficients in each subband as above, then reconstruct each subband independently. In other words, if P_{W_j} is the projection operator onto the space spanned by the columns of Ψ_j, we measure

$$y^j = F_{\Omega_j} P_{W_j} x^0$$

for $j = 1, \ldots, J$, then set w^j to be the solution to

$$\min \ \|w\|_{\ell_1} \quad \text{subject to} \quad F_{\Omega_j} \Psi_j w = y^j.$$

If all of the wavelet subbands of the object we are imaging are appropriately sparse, we will be able to recover the image perfectly.

Finally, we would like to note that this projection onto W_j in the measurement process can be avoided by constructing the wavelet and sampling systems a little more carefully. In [12], a "bi-sinusoidal" measurement system is introduced which complements the orthonormal Meyer wavelet

transform. These bi-sinusoids are an alternative orthobasis to the W_j spanned by Meyer wavelets at a given scale (with perfect mutual incoherence), so sampling in the bi-sinusoidal basis isolates a given wavelet subband automatically.

In the next section, we examine an orthogonal measurement system which allows us to forgo this subband separation all together.

2.2 Noiselet measurements

In [8], a complex "noiselet" system is constructed that is perfectly incoherent with the Haar wavelet representation. If Ψ is an orthonormal system of Haar wavelets, and Φ is the orthogonal noiselet system (renormalized so that $\Phi^*\Phi = nI$), then $U = \Phi\Psi$ has entries of constant magnitude:

$$|U_{k,j}| = 1, \quad \forall k, j \quad \text{which implies} \quad \mu(U) = 1.$$

Just as the canonical basis is maximally incoherent with the Fourier basis, so is the noiselet system with Haar wavelets. Thus if an n-pixel image is S-sparse in the Haar wavelet domain, it can be recovered (with high probability) from $\sim S \log n$ randomly selected noiselet coefficients.

In addition to perfect incoherence with the Haar transform, noiselets have two additional properties that make them ideal for coded image acquisition:

1. The noiselet matrix Φ can be decomposed as a multiscale filterbank. As a result, it can be applied $O(n \log n)$ time.

2. The real and imaginary parts of each noiselet function are binary valued. A noiselet measurement of an image is just an inner product with a sign pattern, which make their implementation in an actual acquisition system easier. (It would be straightforward to use them in the imaging architecture proposed in [26], for example.)

A large-scale numerical example is shown in Figure 2. The $n = 1024^2$ pixel synthetic image in panel (a) is an exact superposition of $S = 25,000$ Haar wavelets[4]. The observation vector y was created from $m = 70,000$ randomly chosen noiselet coefficients (each noiselet coefficient has a real and imaginary part, so there are really $140,000$ real numbers recorded). From y, we are able to recover the image exactly by solving (1.5).

This result is a nice demonstration of the compressed sensing paradigm. A traditional acquisition process would measure all $n \sim 10^6$ pixels, transform into the wavelet domain, and record the S that are important. Many measurements are made, but comparably very few numbers are recorded. Here we take only a fraction of the number of measurements, and are able to find the S active wavelets coefficients without any prior knowledge of their locations.

The measurement process can be adjusted slightly in a practical setting. We know that almost all of the coarse-scale wavelet coefficients will be important (see Figure 2(b)), so we can potentially

---

[4]The image was created in the obvious way: the well-known test image was transformed into the Haar domain, all but the $25,000$ largest Haar coefficients were set to zero, and the result inverse transformed back into the spatial domain.
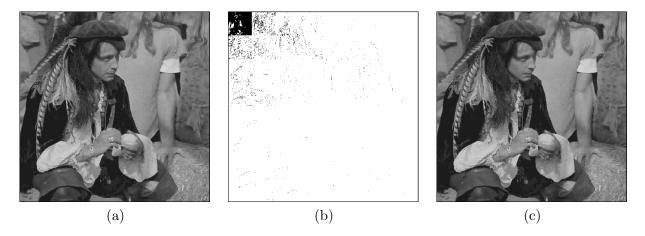
Figure 2: Sparse image recovery from noiselet measurements. (a) Synthetic $n = 1024^2$-pixel image with $S = 25,000$ non-zero Haar wavelet coefficients. (b) Locations (in the wavelet quadtree) of significant wavelet coefficients. (c) Image recovered from $m = 70,000$ complex noiselet measurements. The recovery matches (a) exactly.

reduce the number of measurements needed for perfect recovery by measuring these directly. In fact, if we measure the 128×128 block of coarse wavelet coefficients for the image in Figure 2 directly (equivalent to measuring averages over 8×8 blocks of pixels, $16,384$ measurement total), we are able to recover the image perfectly from an additional $41,808$ complex noiselet measurements (the total number of real numbers recorded is $100,000$).

3 Proofs

3.1 General strategy

The proof of Theorem 1.1 follows the program set forth in [4, 15]. As detailed in these references, the signal x^0 is the unique solution to (1.5) if and only if there exists a *dual vector* $\pi \in \mathbb{R}^n$ with the following properties:

- π is in the row space of U_Ω,

- $\pi(t) = \operatorname{sgn} x^0(t)$ for $t \in T$, and

- $|\pi(t)| < 1$ for $t \in T^c$.

We consider the candidate

$$\pi = U_\Omega^* U_{\Omega T} (U_{\Omega T}^* U_{\Omega T})^{-1} z_0, \tag{3.1}$$

where z_0 is a $|T|$-dimensional vector whose entries are the signs of x^0 on T, and show that under the conditions in the theorem 1) π is well defined (i.e. $U_{\Omega T}^* U_{\Omega T}$ is invertible), and given this 2) $|\pi(t)| < 1$ on T^c (we automatically have that π is in the row space of U_Ω and $\pi(t) = \operatorname{sgn} x(t)$ on T).

10

We want to show that with the support fixed, a dual vector exists with high probability when selecting Ω uniformly at random. Following [4], it is enough to show that the desired properties when Ω is sampled using a Bernoulli model. Suppose Ω_1 of size m is sampled uniformly at random, and Ω_2 is sampled by setting

$$\Omega_2 := \{k : \delta_k = 1\},$$

where here and below $\delta_1, \delta_2, \ldots, \delta_n$ is a sequence of independent identically distributed 0/1 Bernoulli random variables with

$$\mathbf{P}(\delta_k = 1) = m/n.$$

Then

$$\mathbf{P}(\text{Failure}(\Omega_1)) \leq 2\,\mathbf{P}(\text{Failure}(\Omega_2)) \tag{3.2}$$

(see [4] for details). With this established, we will establish the existence of a dual vector for x^0 with high probability for Ω sampled using the Bernoulli model.

The matrix $U^*_{\Omega T} U_{\Omega T}$ is now a random variable, which can be written as

$$U^*_{\Omega T} U_{\Omega T} = \sum_{k=1}^{n} \delta_k u^k \otimes u^k,$$

where the u^k are the row vectors of U_T; $u^k = (U_{t,k})_{t \in T}$.

3.2   Proof of Theorem 1.2

Our first result, which is an analog to a theorem of Rudelson [24, Th. 1], states that if m is large enough, then on average the matrix $m^{-1} U^*_{\Omega T} U_{\Omega T}$ deviates little from the identity.

Theorem 3.1 *Let U be an orthogonal matrix obeying (1.1). Consider a fixed set T and let Ω be a random set sampled using the Bernoulli model. Then*

$$\mathbf{E}\,\|\frac{1}{m} U^*_{\Omega T} U_{\Omega T} - I\| \leq C_R \cdot \frac{\sqrt{\log |T|}}{\sqrt{m}} \max_{1 \leq k \leq n} \|u^k\| \tag{3.3}$$

for some positive constant C_R, provided the right-hand side is less than 1. Since the coherence $\mu(U)$ obeys

$$\max_{1 \leq k \leq n} \|u^k\| \leq \mu(U)\sqrt{|T|},$$

this implies

$$\mathbf{E}\,\|\frac{1}{m} U^*_{\Omega T} U_{\Omega T} - I\| \leq C_R \cdot \mu(U) \frac{\sqrt{|T| \log |T|}}{\sqrt{m}}. \tag{3.4}$$

The probabilistic model is different here than in [24]. The argument, however, is similar.

Proof We are interested in $\mathbf{E}\,\|Y\|$ where Y is the random sum

$$Y = \frac{1}{m} \sum_{k=1}^{n} \delta_k\, u^k \otimes u^k - I.$$

11

Note that since $U^*U = nI$,

$$\mathbf{E}\,Y = \frac{1}{m}\sum_{k=1}^{n}\frac{m}{n}\,u^k \otimes u^k - I = \frac{1}{n}\sum_{k=1}^{n}u^k \otimes u^k - I = 0.$$

We now use a symmetrization technique to bound the expected value of the norm of Y. We let Y' be an independent copy of Y, i.e.

$$Y' = \frac{1}{m}\sum_{k=1}^{n}\delta_k'\,u^k \otimes u^k - I, \tag{3.5}$$

where $\delta_1', \ldots, \delta_n'$ are independent copies of $\delta_1, \ldots, \delta_n$, and write

$$\mathbf{E}\,\|Y\| \le \mathbf{E}\,\|Y - Y'\|,$$

which follows from Jensen's inequality and the law of iterated expectation (also known as Fubini's theorem). Now let $\epsilon_1, \ldots, \epsilon_n$ be a sequence of Bernoulli variables taking values ± 1 with probability $1/2$ (and independent of the sequences δ and δ'). We have

$$\mathbf{E}\,\|Y\| \le \mathbf{E}_{\delta,\delta'}\,\|\frac{1}{m}\sum_{k=1}^{n}(\delta_k - \delta_k')\,u^k \otimes u^k\|$$

$$= \mathbf{E}_\epsilon\,\mathbf{E}_{\delta,\delta'}\,\|\frac{1}{m}\sum_{1\le k\le n}\epsilon_k(\delta_k - \delta_k')\,u^k \otimes u^k\|$$

$$\le 2\,\mathbf{E}_\epsilon\,\mathbf{E}_\delta\,\|\frac{1}{m}\sum_{1\le k\le n}\epsilon_k\delta_k\,u^k \otimes u^k\|;$$

the first equality follows from the symmetry of the random variable $(\delta_k - \delta_k')\,u^k \otimes u^k$ while the last inequality follows from the triangle inequality.

We may know apply Rudelson's powerful lemma [24] which states that

$$\mathbf{E}_\epsilon\,\|\sum_{k=1}^{n}\epsilon_k\delta_k\,u^k \otimes u^k\| \le C_R/4 \cdot \sqrt{\log|T|} \cdot \max_{k:\delta_k=1}\|u^k\| \cdot \sqrt{\|\sum_{k=1}^{n}\delta_k u^k \otimes u^k\|} \tag{3.6}$$

for some universal constant $C_R > 0$ (the notation should make it clear that the left-hand side is only averaged over ϵ). Taking expectation over δ then gives

$$\mathbf{E}\,\|Y\| \le C_R/2 \cdot \frac{\sqrt{\log|T|}}{m} \cdot \max_{1\le k\le n}\|u^k\| \cdot \mathbf{E}\sqrt{\|\sum_{k=1}^{n}\delta_k u^k \otimes u^k\|}$$

$$\le C_R/2 \cdot \frac{\sqrt{\log|T|}}{m} \cdot \max_{1\le k\le n}\|u^k\| \cdot \sqrt{\mathbf{E}\,\|\sum_{k=1}^{n}\delta_k u^k \otimes u^k\|}, \tag{3.7}$$

where the second inequality uses the fact that for a nonnegative random variable Z, $\mathbf{E}\sqrt{Z} \le \sqrt{\mathbf{E}\,Z}$. Observe now that

$$\mathbf{E}\,\|\sum_{k=1}^{n}\delta_k u^k \otimes u^k\| = \mathbf{E}\,\|mY + mI\| \le m(\mathbf{E}\,\|Y\| + 1)$$

and, therefore, (3.7) gives

$$\mathbf{E}\,\|Y\| \le a \cdot \sqrt{\mathbf{E}\,\|Y\| + 1}, \qquad a = C_R/2 \cdot \frac{\sqrt{\log|T|}}{\sqrt{m}} \cdot \max_{1 \le k \le n} \|u^k\|.$$

It then follows that if $a \le 1$,

$$\mathbf{E}\,\|Y\| \le 2a,$$

which concludes the proof of the theorem. ∎

With Theorem 3.1 established, we have a bound on the expected value of $\|m^{-1}U_{\Omega T}^* U_{\Omega T} - I\|$. Theorem 1.2 shows that $m^{-1}U_{\Omega T}^* U_{\Omega T}$ is close to the identity *with high probability*, turning the statement about expectation into a corresponding large deviation result.

The proof of Theorem 1.2 uses remarkable estimates about the large deviations of suprema of sums of independent random variables. Let $Y_1, \ldots, Y_n$ be a sequence of independent random variables taking values in a Banach space and let Z be the supremum defined as

$$Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(Y_i), \tag{3.8}$$

where $\mathcal{F}$ is a countable family of real-valued functions. In a striking paper, Talagrand [27] proved a concentration inequality about Z which is stated below, see also [19][Corollary 7.8].

Theorem 3.2 *Assume that $|f| \le B$ for every f in $\mathcal{F}$, and $\mathbf{E}\,f(Y_i) = 0$ for every f in $\mathcal{F}$ and $i = 1, \ldots, n$. Then for all $t \ge 0$,*

$$\mathbf{P}(|Z - \mathbf{E}\,Z| > t) \le 3\exp\left(-\frac{t}{KB}\log\left(1 + \frac{Bt}{\sigma^2 + B\,\mathbf{E}\,\bar{Z}}\right)\right), \tag{3.9}$$

where $\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \mathbf{E}\,f^2(Y_i)$, $\bar{Z} = \sup_{f \in \mathcal{F}} |\sum_{i=1}^{n} f(Y_i)|$, and K is a numerical constant.

We note that very precise values of the numerical constant K are known and are small, see [22] and [18, 23].

Proof of Theorem 1.2. Set Y to be the matrix $\frac{1}{m}U_{\Omega T}^* U_{\Omega T} - I$ and recall that $\frac{1}{n}\sum_{k=1}^{n} u^k \otimes u^k = I$, which allows to express Y as

$$Y = \sum_{k=1}^{n}\left(\delta_k - \frac{m}{n}\right)\frac{u^k \otimes u^k}{m} := \sum_{k=1}^{n} Y_k,$$

where

$$Y_k := \left(\delta_k - \frac{m}{n}\right)\frac{u^k \otimes u^k}{m}.$$

13

Note that $\mathbf{E}\,Y_k = 0$. We are interested in the spectral norm $\|Y\|$. By definition,

$$\|Y\| = \sup_{f_1,f_2} \langle f_1, Y f_2 \rangle = \sup_{f_1,f_2} \sum_{k=1}^{n} \langle f_1, Y_k f_2 \rangle,$$

where the supremum is over a countable collection of unit vectors. For a fixed pair of unit vectors (f_1, f_2), let $f(Y_k)$ denote the mapping $\langle f_1, Y_k f_2 \rangle$. Since $\mathbf{E}\,f(Y_k) = 0$, we can apply Theorem 3.2 with B obeying

$$|f(Y_k)| \leq \frac{|\langle f_1, u^k \rangle \langle u^k, f_2 \rangle|}{m} \leq \frac{\|u^k\|^2}{m} \leq B, \quad \text{for all } k.$$

As such, we can take $B = \max_{1 \leq k \leq n} \|u^k\|^2 / m$. We now compute

$$\mathbf{E}\,f^2(Y_k) = \frac{m}{n}\left(1 - \frac{m}{n}\right) \frac{|\langle f_1, u^k \rangle \langle u^k, f_2 \rangle|^2}{m^2}$$

$$\leq \frac{m}{n}\left(1 - \frac{m}{n}\right) \frac{\|u^k\|^2}{m^2} |\langle u^k, f_2 \rangle|^2.$$

Since $\sum_{1 \leq k \leq n} |\langle u^k, f_2 \rangle|^2 = n$, we proved that

$$\sum_{1 \leq k \leq n} \mathbf{E}\,f^2(Y_k) \leq \left(1 - \frac{m}{n}\right) \frac{1}{m} \max_{1 \leq k \leq n} \|u^k\|^2 \leq B.$$

In conclusion, with $Z = \|Y\| = \bar{Z}$, Theorem 3.2 shows that

$$\mathbf{P}(|\,\|Y\| - \mathbf{E}\,\|Y\|\,| > t) \ \leq\ 3\exp\left(-\frac{t}{KB}\log\left(1 + \frac{t}{1 + \mathbf{E}\,\|Y\|}\right)\right). \tag{3.10}$$

Take m large enough so that $\mathbf{E}\,\|Y\| \leq 1/4$ in (3.4), and pick $t = 1/4$. Since $B \leq \mu^2(U)|T|/m$, (3.10) gives

$$P(\|Y\| > 1/2) \leq 3e^{-\frac{m}{C_T \mu^2(U)|T|}},$$

for $C_T = 4K/\log(6/5)$. Taking $C_1 = 16C_R$ and $C_2 = C_T$ finishes the proof. $\blacksquare$

3.3 Proof of Theorem 1.1

With Theorem 1.2 established, we know that with high probability the eigenvalues of $U_{\Omega T}^* U_{\Omega T}$ will be tightly controlled — they are all between $m/2$ and $3m/2$. Under these conditions, the inverse of $(U_{\Omega T}^* U_{\Omega T})$ not only exists, but we can guarantee that $\|(U_{\Omega T}^* U_{\Omega T})^{-1}\| \leq 2/m$, a fact which we will use to show $|\pi(t)| < 1$ for $t \in T^c$.

For a particular $t_0 \in T^c$, we can rewrite $\pi(t_0)$ as

$$\pi(t_0) = \langle v^0, (U_{\Omega T}^* U_{\Omega T})^{-1} z \rangle = \langle w^0, z \rangle,$$

where v^0 is the row vector of $U_{\Omega}^* U_{\Omega T}$ with row index t_0, and $w^0 = (U_{\Omega T}^* U_{\Omega T})^{-1} v^0$. The following three lemmas give estimates for the sizes of these vectors. From now on and for simplicity, we drop the dependence on U in $\mu(U)$.

14

Lemma 3.1 *The second moment of $Z_0 := \|v^0\|$ obeys*

$$\mathbf{E}\, Z_0^2 \leq \mu^2\, m\, |T|. \tag{3.11}$$

Proof Set $\lambda_k^0 = u_{k,t_0}$. The vector v^0 is given by

$$v^0 = \sum_{k=1}^n \delta_k\, \lambda_k^0\, u_k = \sum_{k=1}^n (\delta_k - \mathbf{E}\,\delta_k)\, \lambda_k^0\, u^k,$$

where the second equality holds due to the orthogonality of the rows of U: $\sum_{1 \leq k \leq n} \lambda_k^0\, u_{k,t} = \sum_{1 \leq k \leq n} u_{k,t_0}\, u_{k,t} = 0$. We thus can view v^0 as a sum of independent random variables:

$$v^0 = \sum_{k=1}^n Y_k, \qquad Y_k = (\delta_k - m/n)\lambda_k^0 u^k, \tag{3.12}$$

where we note that $\mathbf{E}\, Y_k = 0$. It follows that

$$\mathbf{E}\, Z_0^2 = \sum_k \mathbf{E}\langle Y_k, Y_k\rangle + \sum_{k' \neq k} \mathbf{E}\langle Y_k, Y_{k'}\rangle = \sum_k \mathbf{E}\langle Y_k, Y_k\rangle.$$

Now
$$\mathbf{E}\,\|Y_k\|^2 = \frac{m}{n}\left(1 - \frac{m}{n}\right)|\lambda_k^0|^2\|u^k\|^2 \leq \frac{m}{n}\left(1 - \frac{m}{n}\right)|\lambda_k^0|^2\,\mu^2|T|.$$

Since $\sum_k |\lambda_k^0|^2 = n$, we proved that

$$\mathbf{E}\, Z_0^2 \leq \left(1 - \frac{m}{n}\right)\mu^2\, m\, |T|.$$

This establishes the claim. ∎

The next result shows that the tail of Z_0 exhibits a Gaussian behavior.

Lemma 3.2 *Fix $t_0 \in T^c$ and let $Z_0 = \|v^0\|$. Define $\overline{\sigma}$ as*

$$\overline{\sigma}^2 = \mu^2\, m \cdot \max(1, \mu|T|/\sqrt{m}).$$

Fix $a > 0$ obeying $a \leq (m/\mu^2)^{1/4}$ if $\mu|T|/\sqrt{m} > 1$ and $a \leq (m/\mu^2|T|)^{1/2}$ otherwise. Then

$$P(Z_0 \geq \mu\sqrt{m|T|} + a\overline{\sigma}) \leq e^{-\gamma a^2}, \tag{3.13}$$

for some positive constant $\gamma > 0$.

The proof of this lemma uses the powerful concentration inequality (3.9).

Proof By definition, Z_0 is given by

$$Z_0 = \sup_{\|f\|=1}\, \langle v^0, f\rangle = \sup_{\|f\|=1}\, \sum_{k=1}^n \langle Y_k, f\rangle$$

(and observe $Z_0 = \bar{Z}_0$). For a fixed unit vector f, let $f(Y_k)$ denote the mapping $\langle Y_k, f \rangle$. Since $\mathbf{E} f(Y_k) = 0$, we can apply Theorem 3.2 with B obeying

$$|f(Y_k)| \leq |\lambda_k^0| \, |\langle f, u^k \rangle| \leq |\lambda_k^0| \, \|u^k\| \leq \mu^2 \, |T|^{1/2} := B.$$

Before we do this, we also need bounds on σ^2 and $\mathbf{E} Z_0$. For the latter, we simply use

$$\mathbf{E} Z_0 \leq \sqrt{\mathbf{E} Z_0^2} \leq \mu \sqrt{m \, |T|}. \tag{3.14}$$

For the former

$$\mathbf{E} f^2(Y_k) = \frac{m}{n} \left(1 - \frac{m}{n} \right) |\lambda_k^0|^2 \, |\langle u^k, f \rangle|^2 \leq \frac{m}{n} \left(1 - \frac{m}{n} \right) \mu^2 \, |\langle u^k, f \rangle|^2.$$

Since $\sum_{1 \leq k \leq n} |\langle u^k, f \rangle|^2 = n$, we proved that

$$\sum_{1 \leq k \leq n} \mathbf{E} f^2(Y_k) \leq m \mu^2 \left(1 - \frac{m}{n} \right).$$

In conclusion, Theorem 3.2 shows that

$$\mathbf{P}(|Z_0 - \mathbf{E} Z_0| > t) \ \leq \ 3 \exp \left(-\frac{t}{KB} \log \left(1 + \frac{Bt}{\mu^2 m + B \mu \sqrt{m|T|}} \right) \right). \tag{3.15}$$

Suppose now $\bar{\sigma}^2 = B \mu \sqrt{m|T|} \geq \mu^2 m$, and fix $t = a\bar{\sigma}$. Then it follows from (3.15) that

$$\mathbf{P}(|Z_0 - \mathbf{E} Z_0| > t) \ \leq \ 3 e^{-\gamma a^2},$$

provided that $Bt \leq \bar{\sigma}^2$. The same is true if $\bar{\sigma}^2 = \mu^2 m \geq B \mu \sqrt{m|T|}$ and $Bt \leq \mu^2 m$. We omit the details. The lemma follows from (3.14). ∎

Lemma 3.3 *Let* $w^0 = (U_{\Omega T}^* U_{\Omega T})^{-1} v^0$. *With the same notations and hypotheses as in Lemma 3.2, we have*

$$\mathbf{P} \left(\sup_{t_0 \in T^c} \|w^0\| \geq 2\mu \sqrt{|T|/m} + 2a\bar{\sigma}/m \right) \leq n \, e^{-\gamma a^2} + \mathbf{P}(\|U_{\Omega T}^* U_{\Omega T}\| \leq m/2). \tag{3.16}$$

Proof Let A and B be the events $\{\|U_{\Omega T}^* U_{\Omega T}\| \geq m/2\}$ and $\{\sup_{t_0 \in T^c} \|v^0\| \leq \mu \sqrt{m \, |T|} + a\bar{\sigma}\}$ respectively, and observe that Lemma 3.2 gives $\mathbf{P}(B^c) \leq n e^{-\gamma a^2}$. On the event $A \cap B$

$$\sup_{t_0 \in T^c} \|w^0\| \leq \frac{2}{m} \left(\mu \sqrt{m \, |T|} + a\bar{\sigma} \right)$$

The claim follows. ∎

Lemma 3.4 *Assume that $z(t)$, $t \in T$ is an i.i.d. sequence of symmetric Bernoulli random variables. For each $\lambda > 0$, we have*

$$\mathbf{P}\left(\sup_{t \in T^c} |\pi(t)| > 1\right) \le 2ne^{-1/2\lambda^2} + \mathbf{P}\left(\sup_{t_0 \in T^c} \|w^0\| > \lambda\right). \tag{3.17}$$

Proof The proof is essentially an application of Hoeffding's inequality [2]. Conditioned on the w^0, this inequality states that

$$\mathbf{P}\left(|\langle w^0, z\rangle| > 1 \mid w^0\right) \le 2e^{-\frac{1}{2\|w^0\|^2}}. \tag{3.18}$$

Recall that $\pi(t_0) = \langle w^0, z\rangle$. It then follows that

$$\mathbf{P}\left(\sup_{t_0 \in T^c} |\pi(t_0)| > 1 \mid \sup_{t_0 \in T^c} \|w^0\| \le \lambda\right) \le 2ne^{-\frac{1}{2\lambda^2}},$$

which proves the result. ∎

The pieces are in place to prove Theorem 1.1. Set $\lambda = 2\mu\sqrt{|T|/m} + 2a\bar{\sigma}/m$. Combining Lemmas 3.4 and 3.3, we have for each $a > 0$ obeying the hypothesis of Lemma 3.2

$$\mathbf{P}\left(\sup_{t \in T^c} |\pi(t)| > 1\right) \le 2ne^{-1/2\lambda^2} + ne^{-\gamma a^2} + \mathbf{P}\left(\|(U_{\Omega T}^* U_{\Omega T})\| \le m/2\right).$$

For the second term to be less than δ, we choose a such that

$$a^2 = \gamma^{-1} \log(n/\delta),$$

and assume this value from now on. The first term is less than δ if

$$\frac{1}{\lambda^2} \ge 2\log(2n/\delta). \tag{3.19}$$

Suppose $\mu|T| \ge \sqrt{m}$. The condition in Lemma 3.2 is $a \le (m/\mu^2)^{1/4}$ or equivalently

$$m \ge \mu^2 \gamma^{-2} [\log(n/\delta)]^2,$$

where γ is a numerical constant. In this case, $a\bar{\sigma} \le \mu\sqrt{m|T|}$ which gives

$$\frac{1}{\lambda^2} \ge \frac{1}{16}\frac{m}{\mu^2|T|}. \tag{3.20}$$

Suppose now that $\mu|T| \le \sqrt{m}$. Then if $|T| \ge a^2$, $a\bar{\sigma} \le \mu\sqrt{m|T|}$ which gives again (3.20). On the other hand if $|T| \le a$, $\lambda \le 4a\bar{\sigma}/m$ and

$$\frac{1}{\lambda^2} \ge \frac{1}{16}\frac{m}{a^2\mu^2}.$$

To verify (3.19), it suffices to take m obeying

$$\frac{m}{16\mu^2}\min\left(\frac{1}{|T|}, \frac{1}{a^2}\right) \ge 2\log(2n/\delta).$$

17

This analysis shows that the second term is less than δ if

$$m \geq K_1\,\mu^2\,\max(|T|, \log(n/\delta))\,\log(n/\delta)$$

for some constant K_1. Finally, by Theorem 1.2, the last term will be bounded by δ if

$$m \geq K_2\,\mu^2\,|T|\,\log(n/\delta)$$

for some constant K_2. In conclusion, we proved that there exists a constant K_3 such that the reconstruction is exact with probability at least $1 - \delta$ provided that the number of measurements m obeys

$$m \geq K_3\,\mu^2\,\max(|T|, \log(n/\delta))\,\log(n/\delta).$$

The theorem is proved.

4   Discussion

It is possible that a version of Theorem 1.1 exists that holds for all sign sequences on a set T simultaneously, i.e. we can remove the condition that the signs are chosen uniformly at random. Proving such a theorem with the methods above would require showing that the random vector $w^0 = (U^*_{\Omega T} U_{\Omega T})^{-1} v^0$, where v^0 is as in (3.12), will not be aligned with the fixed sign sequence z. We conjecture that this is indeed true, but proving such a statement seems considerably more involved.

The new large-deviation inequality of Theorem 1.2 can also be used to sharpen results presented in [3] about using ℓ_1 minimization to find the sparsest decomposition of a signal in a union of bases. Consider a signal $f \in \mathbb{R}^n$ that can be written as a sparse superposition of the columns of a dictionary $D = (\Psi_1\ \Psi_2)$ where each Ψ_i is an orthonormal basis. In other words $f = Dx^0$, where $x^0 \in \mathbb{R}^{2n}$ has small support. Given such an f, we attempt to recover x^0 by solving

$$\min_x\ \|x\|_{\ell_1}\quad \text{subject to}\quad Dx = f. \tag{4.1}$$

Combining Theorem 1.2 with the methods used in [3], we can establish that if

$$|\operatorname{supp} x| \leq \text{Const} \cdot \frac{n}{\mu^2(\Psi_1^*\Psi_2) \cdot \log n},$$

then the following will be true with high probability (where the support and signs of x^0 are drawn at random):

1. There is no $x \neq x^0$ with $|\operatorname{supp} x| \leq |\operatorname{supp} x^0|$ with $f = Dx$. That is, x^0 is the sparsest possible decomposition of f.

2. We can recover x^0 from f by solving (4.1).

This is a significant improvement over the bounds presented in [3], which have logarithmic factors of $(\log n)^6$.

References

[1] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. The Johnson-Lindenstrauss lemma meets compressed sensing. Submitted manuscript, June 2006.

[2] S. Boucheron, O. Bousquet, and G. Lugosi. Concentration inequalities. In O. Bousquet and G. Rätsch, editors, *Advanced Lectures in Machine Learning*, pages 208–240. Springer, 2004.

[3] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math.*, 6(2):227–254, 2006.

[4] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, February 2006.

[5] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Applied Math.*, 59(8):1207–1223, 2006.

[6] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, December 2005.

[7] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inform. Theory*, 52(12), December 2006.

[8] R. Coifman, F. Geshwind, and Y. Meyer. Noiselets. *Appl. Comp. Harmonic Analysis*, 10:27–44, 2001.

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

[10] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, 2006.

[11] D. L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. Sci. USA*, 100:2197–2202, 2003.

[12] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[13] D. L. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.

[14] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of R^N bases. *IEEE Trans. Inform. Theory*, 48:2558–2567, 2002.

[15] J. J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory*, 50:1341–1344, 2004.

[16] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49:3320–3325, 2003.

[17] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. Inform. Theory*, 52(9):4036–4048, September 2006.

[18] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.

[19] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.

[20] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgegiete*. Springer-Verlag, Berlin, 1991.

[21] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, second edition, 1999.

[22] P. Massart. About the contants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[23] E. Rio. Une inégalité de Bennett pour les maxima de processes empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002.

[24] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.

[25] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *CISS*, 2006.

[26] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk. A new compressive imaging camera architecture using optical-domain comrpession. In *Proc. SPIE Conference on Computational Imaging IV*, San Jose, CA, January 2006.

[27] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[28] J. Tropp and A. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. Submitted manuscript, April 2005.

[29] J. A. Tropp. Random subdictionaries of general dictionaries. Submitted manuscript, August 2006.