Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 86 (2006) 533–548

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Breakdown of equivalence between the minimal ℓ^1-norm solution and the sparsest solution

Yaakov Tsaig, David L. Donoho*

Department of Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305-4065, USA

Received 11 November 2004; received in revised form 17 May 2005; accepted 30 May 2005
Available online 3 August 2005

Abstract

Finding the sparsest solution to a set of underdetermined linear equations is NP-hard in general. However, recent research has shown that for certain systems of linear equations, the sparsest solution (i.e. the solution with the smallest number of nonzeros), is also the solution with minimal ℓ^1 norm, and so can be found by a computationally tractable method.

For a given n by m matrix Φ defining a system $y = \Phi\alpha$, with $n<m$ making the system underdetermined, this phenomenon holds whenever there exists a 'sufficiently sparse' solution α_0. We quantify the 'sufficient sparsity' condition, defining an *equivalence breakdown point* (EBP): the degree of sparsity of α required to guarantee equivalence to hold; this threshold depends on the matrix Φ.

In this paper we study the size of the EBP for 'typical' matrices with unit norm columns (the uniform spherical ensemble (USE)); Donoho showed that for such matrices Φ, the EBP is at least proportional to n. We distinguish three notions of breakdown point—*global*, *local*, and *individual*—and describe a semi-empirical heuristic for predicting the local EBP at this ensemble. Our heuristic identifies a configuration which can cause breakdown, and predicts the level of sparsity required to avoid that situation. In experiments, our heuristic provides upper and lower bounds bracketing the EBP for 'typical' matrices in the USE. For instance, for an $n \times m$ matrix $\Phi_{n,m}$ with $m = 2n$, our heuristic predicts breakdown of local equivalence when the coefficient vector α has about 30% nonzeros (relative to the reduced dimension n). This figure reliably describes the observed empirical behavior. A rough approximation to the observed breakdown point is provided by the simple formula $0.44 \cdot n/\log(2m/n)$.

There are many matrix ensembles of interest outside the USE; our heuristic may be useful in speeding up empirical studies of breakdown point at such ensembles. Rather than solving numerous linear programming problems *per* n, m combination, at least several for each degree of sparsity, the heuristic suggests to conduct a few experiments to measure the driving term of the heuristic and derive predictive bounds. We tested the applicability of this heuristic to three special ensembles of matrices, including the partial Hadamard ensemble and the partial Fourier ensemble, and found

*Corresponding author. Tel.: +1 650 723 3350; fax: +1 650 725 8977.
E-mail address: donoho@stat.stanford.edu (D.L. Donoho).

doi:10.1016/j.sigpro.2005.05.028

that it accurately predicts the sparsity level at which local equivalence breakdown occurs, which is at a lower level than for the USE. A rough approximation to the prediction is provided by the simple formula $0.65 \cdot n/\log(1+10m/n)$.

Keywords: Basis pursuit; Underdetermined systems of linear equations; Random matrix theory; Linear programming; Overcomplete systems; Sparse representations; Random signs matrix ensemble; Partial Fourier matrix ensemble; Partial Hadamard matrix ensemble

1. Introduction

Traditionally, many practitioners in signal and image processing observed strict basis monogamy; they chose a fixed representation—Fourier, Gabor, Hadamard, Wavelet—and stuck with it, come what may. Recently, perhaps mirroring a loosening of traditional attitudes in other matters, some practitioners experimented with serial basis monogamy—choosing a new basis with each project [1]. Finally, coinciding with 'postmodern' attitudes in general culture, some researchers call for a rejection of monogamy altogether: advocating to represent signals and images using several different bases at once [1–3].

This approach faces an obvious immediate difficulty, stemming from traditional notions of linear algebra. If we have a signal vector y containing n entries and B different orthonormal bases, represented by $n \times n$ matrices $\Phi_1, ..., \Phi_B$, the matrix $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_B]$ defines a linear system

$$y = \Phi\alpha; \tag{1.1}$$

any α solving this system of equations gives a representation of y. Since Φ is n by Bn, with $B > 1$, there are many such α's. By traditional attitudes—for example, as symbolized by college-level instruction in linear algebra—this indeterminacy is problematic.

Recently, researchers have embraced the variety of solutions, seeing it not as an obstacle, but an advantage, allowing to select from among many solutions the sparsest one. Here sparsity is considered an important objective per se, allowing simpler description, more compression, or more noise removal than a non-sparse solution.

In general the goal of finding sparse solutions to underdetermined linear systems is known to be NP-hard; it contains classic combinatorial optimization problems as subcases [4]. Against this, it has been recently discovered that, at least for many matrices Φ of interest, if there is a sufficiently sparse solution of (1.1), it can be found by tractable convex optimization.

1.1. Equivalence phenomena

Let us formalize matters. Let Φ be an $n \times m$ matrix—not necessarily of the kind described above—whose columns are scaled to unit norm in ℓ^2_n, and with $m > n$ so that Eq. (1.1) is underdetermined. For a vector $y \in \mathbf{R}^n$ we seek the sparsest possible representation of y using columns of Φ; this solves the optimization problem

$$(P_0) \quad \min \|\alpha\|_0 \quad \text{subject to } \Phi\alpha = y,$$

where the ℓ^0 quasi-norm is defined as

$$\|\alpha\|_0 = \#\{i : \alpha_i \neq 0, 1 \leqslant i \leqslant m\},$$

i.e, the ℓ^0 norm measures the number of nonzero elements in α.

In general solving (P_0) involves combinatorial optimization and is considered intractable. Instead, the *basis pursuit* approach [3] considers the minimal ℓ^1-norm representation:

$$(P_1) \quad \min \|\alpha\|_1 \quad \text{subject to } \Phi\alpha = y.$$

This poses a convex optimization problem, and in principle is more tractable than (P_0). In fact, when working in the real setting, (P_1) can be written as a linear program and solved as such [3].

The equivalence phenomenon is the following: there is a constant EBP(Φ) depending only on Φ so that whenever y permits a decomposition $y = \Phi\alpha_0$ with α_0 having fewer than EBP (equivalence breakdown point) nonzeros, then α_0 is the unique solution to both (P_0) and (P_1). This phenomenon has been studied by numerous researchers in recent years [5–14]. One stream of research [5–11] showed, among other things, that if the Gram matrix $\Phi^T\Phi$ has all its off-diagonal entries smaller than M, then $\text{EBP} \geqslant (1 + M^{-1})/2$. Hence, even though Φ cannot have orthogonal columns (as $m > n$), if they are all near-orthogonal, EBP can be substantial; special examples can be given where $M = 1/\sqrt{n}$ while $m = n^2$ [15]; in a measure-theoretic sense, the case where $M \approx \sqrt{2\log(mn)/n}(1 + o(1))$ is prevalent [6], and so for large n and m, 'most' Φ have

$$\text{EBP}(\Phi_{n,m}) \geqslant \sqrt{\frac{n}{2\log(mn)}}(1 + o(1)). \tag{1.2}$$

Thus, for many matrices Φ, if there are no more than roughly $O(\sqrt{n})$ nonzeros, solutions to (P_0) and (P_1) are equivalent.

Candès et al. [13] and Candés and Tao [16] started a new direction, by considering large random matrices from the partial Fourier ensemble. Starting from an m by m Fourier matrix, select n rows at random, $n < m$. They showed that for some $\alpha > 0$, one typically obtains a matrix Φ with

$$\text{EBP}(\Phi) \geqslant \alpha \frac{n}{\log(m)}.$$

This made it clear that results far surpassing the (1.2) case ought to be possible.

Donoho then considered matrices chosen at random from the uniform spherical ensemble (USE), i.e. collection of all matrices Φ with columns of unit length. Equivalently, for such matrices the columns are chosen iid at random from the unit sphere $\mathbf{S}^{n-1}$ in R^n. He was able to show that, if we consider a sequence (m_n) of problem sizes with $m_n \leqslant An$ for some $A > 0$, there is a constant $\rho^*(A) > 0$ so that for random Φ_{n,m_n}

$$\text{Prob}\{n^{-1}\text{EBP}(\Phi_{n,m}) \geqslant \rho^*(A)\} \to 1, \quad n \to \infty.$$

In words, for most large matrices, $\text{EBP}(\Phi_{n,m})$ is proportional to n. While it is possible to calculate an explicit lower bound for ρ^* based on the proof in [14], it is pessimistically small. In the following, we write simply EBP to denote $\text{EBP}(\Phi_{n,m})$ where the meaning is clear from context.

1.2. Individual and local equivalence

The notion of equivalence behind the EBP is both strong and difficult to quantify precisely. In principle, to verify that $\text{EBP} \geqslant k$ would seem to require solving a huge number of linear programs. For each possible support set of size k and each pattern of k signs on that support, we would have to run a linear program to check whether a corresponding instance of (P_1) had a unique solution, equal to the known sparse solution.

However, this is more than one really wants to know. In particular, one might be interested in one of the following two questions:

- For a specific sparse α solving $y = \Phi\alpha$, is this the sparsest representation and the unique ℓ^1 solution?
- For a specific support set I, will every y generated as $y = \Phi\alpha$ by a vector α_0 supported in I, lead to a unique ℓ^1 solution by that α_0?

We call the first notion *individual equivalence* and the second one, *local equivalence*. Clearly, local equivalence at I implies individual equivalence at every α supported in I; just as clearly, $\text{EBP} > |I|$ implies local equivalence at I.

We are interested in what follows, only in matrices Φ which have columns in general position, or equivalently, full Kruskal rank [17]. (A referee asked that we emphasize that not all matrices have columns in general position; thus $\Phi = [IF]$ where I is the $n \times n$ identity and F the $n \times n$ Fourier matrix, will not have columns in general position when m is non-prime.)

For matrices with columns in general position, if there is any solution α with fewer than $n/2$ nonzeros, then that solution is the unique sparsest solution [6]. It follows that, if we have local equivalence at I with $|I| < n/2$, then the solutions of (P_1) and (P_0) are the same for every α supported in I [14].

Definition 1.1. We say that *local equivalence* holds for $\Phi_{n,m}$ at a specific subset I if, for all vectors $y = \Phi\alpha_0$ with $supp(\alpha_0) \in I$, the minimum ℓ^1 solution to (P_1) has $\alpha_1 = \alpha_0$.

As we have suggested, for any particular signal of interest, we are only interested in whether local equivalence holds for that signal. Global equivalence, which is stronger, provides a guarantee that goes beyond what we need, as it refers to other I than the one actually occurring.

We now turn to study the size $|I|$ for which local equivalence can hold. We focus on the 'typical' Φ chosen from the space of n by m matrices with unit-length columns. Equivalently, we consider a random dictionary Φ with columns chosen iid from the unit sphere. For this dictionary, the probability of the event 'local equivalence holds for I' depends only on $|I|$.

Definition 1.2. Let $E_{k,n} = \{$local equivalence *holdsat* $I = \{1, ..., k\}\}$. The events $E_{k,n}$ are decreasing with increasing k. The *Local Equivalence Breakdown Point* $\mathrm{LEBP}(\Phi_{n,m})$ is the smallest k for which event $E^{c}_{k,n}$ occurs.

It holds trivially that $\mathrm{EBP}(\Phi_{n,m}) \leqslant \mathrm{LEBP}(\Phi_{n,m})$.

We note that the event $\{\mathrm{LEBP} > k\}$ could be verified computationally with far less work than EBP, but that it would still require exponentially many different linear programs (one for each sign sequence on $1, \ldots, k$).

1.3. Empirical breakdown results

In a manner similar to that in which we defined the notion of local equivalence, we can define individual equivalence.

Definition 1.3. We say that *individual equivalence* holds at a specific instance (y, Φ), if α_0 is the sparsest solution to $y = \Phi\alpha$ and if the minimum ℓ^1 solution to $y = \Phi\alpha$ is unique equal to α_0.

This of course is directly accessible to computation. We now give experimental evidence about the size of $|I|$ for which EBP of (P_0) and (P_1) observed in practice; the experiment is similar to earlier experiments in [5,13].

For n, m given, we generate an $n \times m$ matrix Φ, with columns iid uniform on $\mathbf{S}^{n-1}$. We create a sparse coefficient vector α_0 consisting of k nonzeros. Specifically, we draw k values at random from a uniform distribution on $(0, 1)$ and place them on a random support set of cardinality k. We then compute $y = \Phi\alpha_0$ and solve (P_1) for α. We performed this experiment 50 times for different values of k, and calculated the percentage of 'perfect' reconstructions, i.e. reconstructions where $\alpha = \alpha_0$ to machine precision. We repeated the experiment at each k, n, m combination.

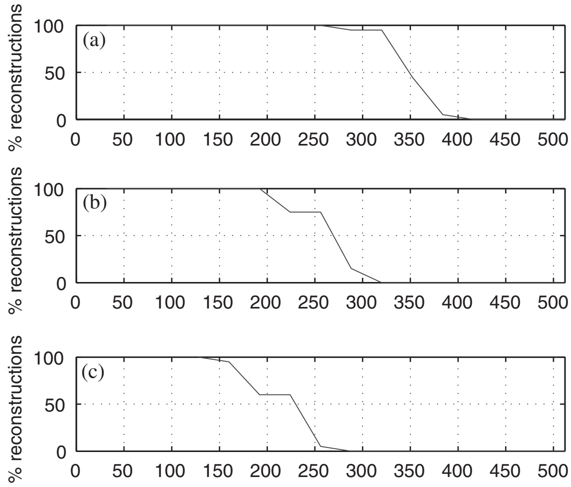

Fig. 1. Individual equivalence for the uniform spherical ensemble. Panels (a)–(c) show percentage of perfect reconstructions versus the size of the support interval I for $n = 1024, m = 2048, 3072$ and 4096, respectively.

Fig. 1 has results. Panels (a)–(c) display the percentage of perfect reconstructions versus k—the size of the interval I—for $n = 1024$ and $m = 2048, 3072$ and 4096, respectively. We observe that even for coefficient vectors with a substantial number of nonzeros relative to n, the ℓ^1-norm solution recovers the sparse solution correctly. In addition, we note that the empirical breakdown point decreases as $A = m/n$ increases.

It seems clear from the above empirical results—and others discussed below—that, for a given $A = m/n$ there is $\rho_I(A)$ so that $\|\alpha_0\|_0 \approx \rho_I n$ marks a 'phase transition' between absence and prevalence of individual equivalence.

1.4. Performance prediction

In this paper, we describe a heuristic to derive bounds $\rho_{\mathrm{L}^\pm}(A)$ with $n\rho_{\mathrm{L}^-}(m/n) \leqslant \mathrm{LEBP} \leqslant n\rho_{\mathrm{L}^+}(m/n)$. We discuss the idea behind the heuristic, which is to propose a problematic 'bad guy' and to propose circumstances under which the 'bad guy' actually causes breakdown.

As we will see, our heuristic supplies bounds which agree with the observable empirical behavior of solutions to (P_1). For instance, for a matrix $\Phi_{n,m}$ drawn from the USE, with $m = 2n$, we get that $\rho_{L^+}(2) \approx 0.33$. In other words, breakdown of local equivalence is predicted when the coefficient vector α has a fraction $\frac{1}{6}$ nonzeros. Indeed, this figure reliably reflects the observed empirical behavior; see Fig. 1(a).

The heuristic is semi-empirical. It requires as input a function v_0 describing the 'scaling properties' of the value of (P_1) for random problems *where there is no a priori sparse representation*. This function gives the median value of (P_1) across an ensemble of randomly generated instances (y, Φ) with different problem sizes n, m, in which by construction the left-hand side y is pure 'noise'. This function, which is of independent interest, can be approximated by solving a series of linear programs and then fitting a curve to the results. For example, we find that a simple form $v_0(m/n) \approx 0.78/\sqrt{\log(m/n)}$ fits well to the USE. Armed with this approximation, the heuristic proposes simple upper and lower bounds for the breakdown point, approximately taking the form:

$$\rho_{L^\pm}(m/n) \approx c_\pm / \log(2m/n), \quad c_+ = 0.44, \quad c_- = 0.34.$$

1.5. Applications

We propose that the heuristic be used to efficiently predict breakdown performances in *other* matrix ensembles. Indeed, the heuristic is derived here within the framework of the USE, where it has a rigorous motivation, but empirically it seems useful in many other ensembles of matrices.

Several examples of such ensembles will be discussed here, and many other can be envisioned, particularly in the context of compressed sensing [18,19]. In such settings m is fixed at the dimension of an object to be sensed, n is variable and we need to determine just how big n should be in order to get good results. We are thus confronted with the problem of determining 'how many' nonzeros a certain matrix ensemble can cope with before the ℓ^1 reconstruction fails.

From a naïve perspective, to determine breakdown properties would seem to require, for each m, n operating point, a large number of experiments trying many different sparsities in order to see at what sparsity level equivalence breaks down. Then it would seem necessary to try many m, n pairs to obtain the desired performance characteristics for a given known level of sparsity. In contrast, the heuristic allows to predict behavior using a relatively small number of experiments to determine the typical value $\hat{v}_0(m/n)$ of randomly generated ℓ^1 minimization problems. Relatively few m, n pairs are necessary to determine the form of the needed scaling and there is no need to cycle through large numbers of vectors of different sparsity.

We illustrate this principle here in the case of two random matrix ensembles seemingly quite different than the USE: the partial Fourier and partial Hadamard ensembles. We verify that the function v_0 behaves differently in these ensembles than for the USE and we obtain from this different predictions which are lower than the USE case, but still accurate for the case at hand. We also note that the two ensembles, while both different from the USE, behave similarly to each other, at least concerning the scaling relation $v_0(m/n)$.

Another application of the heuristic is to support further theoretical studies. Our observed estimate v_0 suggests that a precise scaling relation may hold, giving $\rho_{L^+}(m/n) \sim 0.44/\log(2m/n)$ at the USE. This concrete prediction is testable theoretically, and the elements of the heuristic formula seem theoretically tractable, although this journal does not seem to be the suitable place for detailed mathematical analysis. Perhaps this work will inspire further theoretical investigations.

1.6. Contents

The rest of this paper is organized as follows. Section 2 describes our heuristic formally, and gives a numerical example of the ingredients that it contains. Section 3 considers the key semi-empirical fact used in the heuristic, the behavior of the value of (P_1) on random non-sparse signals, as a function of m/n. Section 4 reviews evidence showing that our heuristic accurately predicts LEBP. Section 5 considers generalizations to other matrix ensembles, namely the random signs, partial Fourier and partial Hadamard ensembles, finding surprising similarities and differences. Section 6 has concluding remarks.

2. The heuristic

We describe a heuristic which assumes that a specific 'failure mode' is responsible for breakdown of equivalence. If that is the case, then a convenient formula allows to derive a predicted range for the breakdown point.

2.1. Form of the upper bound

We start by defining a certain random ℓ^1 optimization problem. Let x be random and uniformly distributed on $\mathbf{S}^{n-1}$ and consider the problem

$$(RP_1(n,m)) \qquad \min \|\alpha\|_1 \quad \text{subject to } \Phi\alpha = x.$$

Here Φ is, as usual, drawn from the USE. Define the random variable $V_{n,m} = \text{val}(RP_1(n,m))$, i.e, the value of the objective function for the problem instance $(RP_1(n,m))$. In [14], it was shown that, for $m \leqslant An$, there is a constant $\eta(A)$ so that, with overwhelming probability for large n,

$$V_{n,m} \geqslant \eta(A)\sqrt{n}.$$

Define the median

$$v(n,m) = \text{med}\{V_{n,m}\};$$

for the median we also have

$$v(n,An) \geqslant \eta(A)\sqrt{n}.$$

There is numerical evidence, described below, that

$$v(n,An) \cdot n^{-1/2} \to v_0(A), \quad n \to \infty, \tag{2.1}$$

where v_0 is a decreasing function of A. In fact, from [14,18] it is clear that, at the USE

$$v_0(A) \asymp 1/\sqrt{\log(A)}, \quad A \to \infty.$$

Under the hypothesis that (2.1) holds, one has the following:

Upper Bound for Breakdown of Local Equivalence. Let $\rho_{L^+} = \rho_{L^+}(A)$ *solve the equation*

$$\frac{\sqrt{\rho}}{(1-\sqrt{\rho})} = \sqrt{\frac{\pi}{2}}\, v_0(A-\rho). \tag{2.2}$$

Fix $\varepsilon > 0$; for $n > n_0(\varepsilon)$ we anticipate that, with probability exceeding $1-\varepsilon$,

$$\text{LEBP}_{n,m}/n \leqslant \rho_{L^+}(m/n) + \varepsilon.$$

2.2. Derivation of the upper bound

Let $I = \{1,\dots,k\} \subset \{1,\dots,m\}$, $n/2 > k > n(\rho_{L^+} + \varepsilon)$. Suppose also that $m-k > n$.

With high probability, we will be able to construct a vector α_0 supported on the fixed support set I of cardinality k so that the convex optimization problem (P_1) generated by $y = \Phi\alpha_0$ does not have its smallest objective value at α_0. As $n/2 > k$, α_0 is the unique sparsest representation of $y = \Phi\alpha_0$ and this gives an example where ℓ^1–ℓ^0 equivalence breaks down. The analysis exhibits a perturbation of α_0 overwhelmingly likely to increase the objective value.

For the given support set I, denote by Φ_I the $n \times |I|$ matrix containing just the columns of Φ belonging to subset I (similarly for Φ_{I^c}). Let $e_{\min}$ denote the right singular vector of Φ_I associated to the smallest nonzero singular value. Because the columns of Φ_I are uniform on the sphere, $e_{\min}$ is uniformly distributed on the sphere S^{k-1} and so all its entries are nonzero with probability one. Let α_0 be an m-vector which is zero outside the subset I and which on the subset I has entries proportional to the *negatives* of the corresponding entries of $e_{\min}$. In particular, outside an event of probability 0,

$$\alpha_0(i)e_{\min}(i) < 0, \quad i \in I;$$

the two vectors have opposite sign patterns. Consider the partial perturbation δ_I given by the eigenvector $e_{\min}$ of $\Phi_I^{\mathrm{T}}\Phi_I$ with smallest eigenvalue. We now complete this to an m-vector δ.

Define the matrix Φ_{I^c} containing just the columns of Φ belonging to subset I^c outside the support set I. Suppose that $m - |I| \geqslant n$, outside a set of probability 0, so that the columns of Φ_{I^c} span $\mathbf{R}^n$. Construct the m-long vector $\delta = (\delta_I, \delta_{I^c})$, where the component δ_{I^c} off the support of I is a special vector obeying

$$\Phi_I\delta_I = -\Phi_{I^c}\delta_{I^c}. \tag{2.3}$$

Such vectors exist because the columns of Φ_{I^c} span $\mathbf{R}^n$ $(m-k > n)$. Note that the constructed perturbation direction δ obeys $\Phi\delta = 0$.

Our vector δ_{I^c} is any solution of the linear program

$$\min \|d\|_1 \quad \text{subject to } \Phi_{I^c} d = -\Phi_I \delta_I, \tag{2.4}$$

where the minimum is over vectors $d \in R^{m-k}$ having entries associated with I^c. Any solution will satisfy (2.3), and, among all such vectors, have the smallest attainable ℓ^1 norm. This turns out to make δ the most clever perturbation for our purposes.

Let $y = \Phi\alpha_0$. We will show that α_0 is not the minimizer of the objective for (P_1). Consider the alternate representation

$$y = \Phi(\alpha_0 + t \cdot \delta),$$

where the perturbation direction was constructed to obey $\Phi\delta = 0$. It will turn out that for small enough $t>0$, $\|\alpha_0 + t\delta\|_1 < \|\alpha\|_1$, i.e. the perturbation decreases the objective value. Hence α_0 is not the solution of (P_1), even though it is sparsely supported.

For all sufficiently small $t>0$,

$$\|\alpha_0 + t \cdot \delta\|_1 - \|\alpha_0\|_1 = t \sum_{i \in I} \mathrm{sgn}(\alpha_0(i))\delta_I(i) + t \sum_{i \in I^c} |\delta_I(i)|.$$

Now by construction $\delta_I(i)\ \mathrm{sgn}(\alpha_0(i)) = -|\delta_I(i)|$, $i \in I$. Hence for sufficiently small $t>0$,

$$\|\alpha_0 + t \cdot \delta\|_1 - \|\alpha_0\|_1 = t(\|\delta_{I^c}\|_1 - \|\delta_I\|_1).$$

Hence, if we can show that

$$\|\delta_I\|_1 > \|\delta_{I^c}\|_1, \tag{2.5}$$

then the perturbation $\delta \neq 0$ reduces the ℓ^1 norm. In words, the perturbation reduces the ℓ^1 norm more *on* the support of α_0 than it increases the norm *off* the support of α_0, so it reduces the ℓ^1 norm overall.

We now derive conditions under which (2.5) must hold. $e_{\min}$ is a uniform random point on $\mathbf{S}^{k-1}$, since Φ_I consists of $|I|$ points from the USE, and the distribution of the eigenvector inherits the rotational symmetry. It is an elementary fact that for such a uniform random point,

$$\|\delta_I\|_1 = \sqrt{\frac{2}{\pi}} \sqrt{|I|} \|\delta_I\|_2 (1 + o_P(1)), \tag{2.6}$$

where $o_P(1)$ denotes negligible order in probability [20]. (Indeed a uniform random point of S^{k-1} can be represented as a vector of standard Gaussian iid $N(0,1)$'s that is then rescaled to lie on the unit sphere. The identity (2.6) is transparent to rescaling, while for a Gaussian vector the counterpart of (2.6) is standard, amounting to

$$E|Z| = \sqrt{2/\pi}, \quad Z \sim N(0,1),$$

and the law of large numbers. The random point δ_I generates $v_I = \Phi_I \delta_I$ with

$$\|v_I\|_2 = (\delta_I^{\mathrm{T}} \Phi_I^{\mathrm{T}} \Phi_I \delta_I)^{1/2} = \lambda_{\min}^{1/2} \|\delta_I\|_2. \tag{2.7}$$

We now estimate the size of $\lambda_{\min}$. Using again the fact that a random uniform point on the sphere is, up to a random scaling by a χ_n random variable, a Gaussian iid vector, one can see immediately as in [14,21] that $\lambda_{\min}$ is essentially the minimal eigenvalue of a Wishart matrix, which has been well-studied. Let $\rho = |I|/n$, we have [22]

$$\lambda_{\min} = (1 - \rho^{1/2})^2 \cdot (1 + o_P(1)).$$

We now estimate the size of $\|\delta_I\|_1$. Note that v_I is a random point on a scaled copy of $\mathbf{S}^{n-1}$, and is stochastically independent of ϕ_i for $i \in I^c$. We see that the defining linear program (2.4) is in fact a scaled version of $RP_1(n, m - |I|)$ and therefore has the same probability distribution as $V_{n,m-|I|} \cdot \|v_I\|_2$. By a standard concentration of measure argument,

$$|V_{n,m} - v(n,m)| = o_P(\sqrt{n}),$$

so we can replace $V_{n,m}$ by $v(n,m)$ for our purposes.

Now if $\rho > \rho_{L^+}$, then by Eq. (2.2) we have

$$\sqrt{\frac{2}{\pi}} \sqrt{\rho} > v_0(A - \rho) \cdot (1 - \sqrt{\rho})$$

and consequently, Eqs. (2.6) and (2.7) give

$$\begin{aligned} \|\delta_I\|_1 &\sim \sqrt{\frac{2}{\pi}} \sqrt{|I|} \|\delta_I\|_2 = \sqrt{\frac{2}{\pi}} \sqrt{\rho n} \|\delta_I\|_2 \\ &> \sqrt{n} \cdot v_0(A - \rho) \cdot (1 - \sqrt{\rho}) \|\delta_I\|_2 \\ &\sim v(n, m - |I|) \|v_I\|_2 \sim \|\delta_{I^c}\|_1. \end{aligned}$$

Hence, for our specific choice of support and specific perturbation,

$$\|\delta_I\|_1 > \|\delta_{I^c}\|_1. \tag{2.8}$$

Thus a small perturbation in the direction of δ can reduce the ℓ^1 norm below that of α_0. As α_0 is the sparsest representation, ℓ^1–ℓ^0 equivalence breaks down.

Our derivation sketches the following conditional result—the proof of which can easily be made rigorous.

Theorem 1. *Suppose that* (2.1) *holds in the following sense*:

$$v(n, An) \to_P v_0(A), \quad n \to \infty, \tag{2.9}$$

with $v_0(A)$ *a positive continuous decreasing function of A. Fix* $\varepsilon > 0$. *The following is true for* $n > n_0(\varepsilon)$. *Consider a subset I chosen independently of* Φ *obeying* $|I| \geqslant (\rho_{L^+} + \varepsilon)n$. *With probability at least* $1 - \varepsilon$, *local equivalence fails at this I.*

This suggests that it would be interesting to prove (2.9).

2.3. Numerical example

To illustrate the upper bound part of our heuristic, we conducted experiments comparing the predicted theoretical bounds and the empirical behavior of some of the measures appearing in the derivation. In each of these experiments, we fixed n and $m = A \cdot n$, and generated a random dictionary $\Phi_{n,m}$ from the USE. Letting k be the size of the support I, we constructed the Gram minor $\Phi_I^T \Phi_I$ and computed its smallest eigenvalue $\lambda_{\min}$ and associated eigenvector $e_{\min}$. We let $\delta_I = e_{\min}$. For clarity of illustration, we normalized δ_I to have unit 1-norm. We then computed $v_I = \Phi \delta_I$ and solved Eq. (2.4). The result, $\|\delta_{I^c}\|_1$, appears in Fig. 2 as a function of ρ, along with the heuristic prediction, for the case $n = 1024, m = 2n$.

Our heuristic seems to give a fairly tight bound on the actual behavior of $\|\delta_{I^c}\|_1$. We observe that, since $\|\delta_I\|_1 = 1$, the breakdown point is the intersection of the curve with the horizontal line 1, which is indeed 0.33 for $A = 2$. Our prediction is of course dependent on our assumed behavior for the limit function $v_0(A)$; see Section 3.

2.4. The lower bound

There is an convenient quasi-lower bound which, while seemingly accurate and computationally tractable, is non-rigorous. (As we will see, a rigorous bound would involve some difficult-to-compute quantities.)

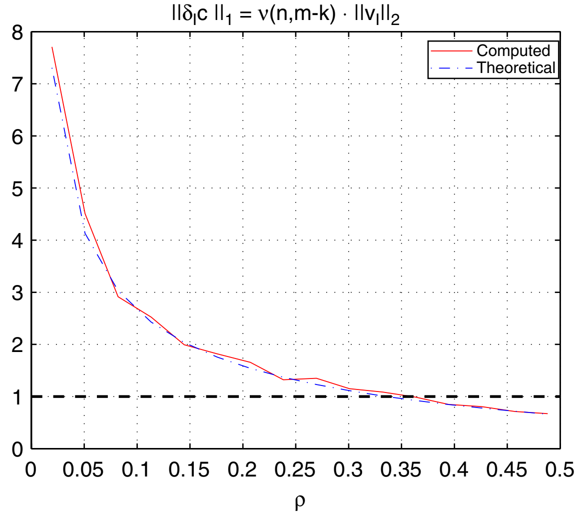

Fig. 2. Illustration of the LEBP heuristic. The solid line shows the empirical bound, and the dash-dotted line shows the theoretical prediction for $n = 1024, m = 2048$.

Quasi-'lower bound' for breakdown of local equivalence. Let $\rho_{L^-} = \rho_{L^-}(A)$ *solve the equation*

$$\frac{\sqrt{\rho}}{(1 - \sqrt{\rho})} = v_0(A - \rho). \tag{2.10}$$

Suppose $m_n/n \to A$ *as* $n \to \infty$, *we anticipate that* $\mathrm{LEBP}_{n,m_n}/n \geqslant \rho_{L^-}(m_n/n) - \varepsilon$.

Under our heuristic, we predict the behavior of LEBP simply by fitting a curve $\hat{v}_0(A)$ summarizing behavior of $v(n, An)$, and plugging that into (2.10). We keep in mind here that the upper bound is on stronger foundation than the lower bound. In the Appendix we give an explanation showing that a rigorously supportable lower bound is smaller than this bound by about a factor $(1 - \sqrt{\rho})$ (a weaker lower bound has been rigorously proven in [14]).

3. Behavior of $v(n, An)$

We now directly study the limit behavior of $v(n, An)/\sqrt{n}$ as a function of A. Our empirical evidence suggests that this indeed tends to a function $v_0(A)$ not depending on n. This comes from the following simulation study. For n, m

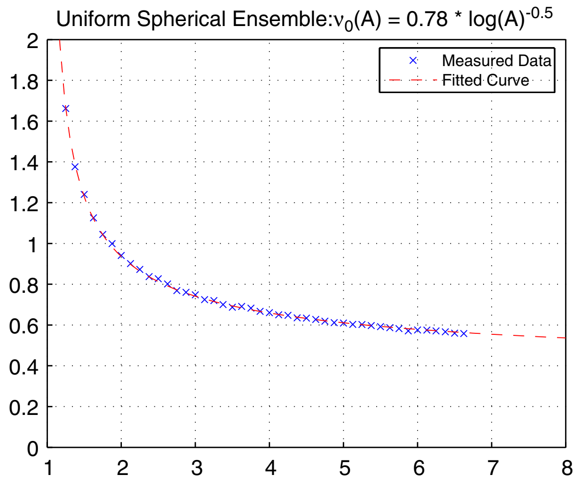

Fig. 3. Empirical estimate of the limit function $v_0(A)$ for the uniform spherical ensemble.

given, we drew a dictionary $\Phi_{n,m}$ at random from the USE, and computed $V_{n,m}$, the value of $RP_1(n,m)$. We repeated this experiment 20 times at each specific (n,m) pair, taking the median at each instance. For the purposes of this simulation, we fixed $n = 1024$ and varied m in the range $[1024, \ldots, 6784]$. To our empirical results we fitted a decaying log-power law of the form $v_0(A) = C \cdot \log(A)^{-\gamma}$. The results are illustrated in Fig. 3. In our experiment, a least-squares line fit on the logarithmic scale resulted in the estimate

$$v_0(A) \approx 0.78 \cdot \log(A)^{-0.5}.$$

As clearly evident in Fig. 3, our model accurately reflects the behavior of the measured data. We note, however, that the range of A we considered in our simulations is rather limited.

4. Predictivity

Based on the estimate for $v_0(A)$ we obtained in the previous section, our heuristic allows us to predict the LEBP bounds ρ_{L^+}, ρ_{L^-} for given values of the overcompleteness factor A. Indeed, computing these bounds amounts to solving the non-linear Eq. (2.2) or (2.10). We feel it would be instructive to compare the heuristic predictions with empirical observations. To that end, we conducted the following simulation study. For n,m given, we generated a dictionary $\Phi_{n,m}$ at random from the USE. In addition, we created a sparse random coefficient vector α_0 consisting of k nonzeros, in a manner similar to that describes in Section 1.3. We then computed $y = \Phi\alpha$, and solved (P_1) for α. We repeated this experiment 50 times for each value of k and calculated the percentage of 'perfect' reconstructions, i.e. reconstructions where $\alpha = \alpha_0$ to machine precision. For each (n,m) pair, we considered a range of values for k, and computed the *Empirical Breakdown Point* to be the smallest value of k for which we get 100% accurate reconstructions.

The results of this simulation study are presented in Table 1, which compares predicted values of ρ_{L^+} and ρ_{L^-}, obtained using our heuristic, with observed empirical values of the breakdown point. We note that the heuristic values for $\rho_{L^+}(A), \rho_{L^-}(A)$ indeed bracket the breakdown points observed in our simulations, despite the modest data sizes utilized.

To gain further insight about the breakdown phenomenon for a wider range of A, we computed $\rho_{L^+}(A)$ and $\rho_{L^-}(A)$ for A in the range $[1, 100]$ and fitted each with a curve of the form $C_1/\log(C_2A)$. The least-squares fit resulted in the estimates

$$\rho_{L^+}(A) = 0.44/\log(2A),$$

$$\rho_{L^-}(A) = 0.34/\log(2A).$$

These curves are depicted in Fig. 4, alongside the computed empirical values. This figure can be

Table 1
Predicted versus empirical breakdown point for the uniform spherical ensemble. ρ_{L^+} and ρ_{L^-} are the heuristic bounds defined by Eqs. (2.2) and (2.10), respectively

n	m	A	ρ_{L^+}	ρ_{L^-}	Empirical LEBP
512	1024	2	0.33	0.26	0.27
512	1536	3	0.24	0.18	0.22
512	2048	4	0.2	0.16	0.18
512	2560	5	0.19	0.14	0.16
1024	2048	2	0.33	0.26	0.26
1024	3072	3	0.24	0.18	0.2
1024	4096	4	0.19	0.14	0.16

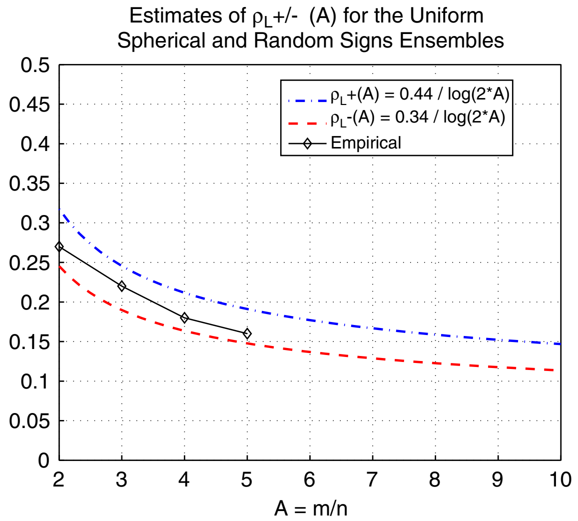

Fig. 4. Estimates of heuristic LEBP bounds. Solid curve shows empirical breakdown values, dashed curve shows ρ_{L^-}, dash-dotted curve shows ρ_{L^+}. The region below the ρ_{L^-} curve is where perfect reconstruction is predicted by the heuristic.

interpreted as displaying a 'phase transition' phenomenon; the region below the ρ_{L^-} curve is where the heuristic predicts that the ℓ^0–ℓ^1 equivalence holds, whereas the region above the ρ_{L^+} curve is the 'failure' region.

5. Universality

Motivated by the apparent success of our heuristic, we now turn to ask if it would work for more varied dictionaries, drawn from different matrix ensembles. We considered three ensembles:

- *Random signs ensemble*. Matrices in this ensemble have entries $\pm 1/\sqrt{n}$ with signs chosen independently and both signs equally likely. This ensemble is known to be effective in various geometric problems associated with underdetermined systems, through work of Kashin [23], followed by Garnaev and Gluskin [24].
- *Partial Hadamard ensemble*. Matrices in this ensemble are obtained by taking an m by m Hadamard matrix and sampling n rows at random. Such matrices are known to be important for various extremal problems in experimental design [25].
- *Partial Fourier ensemble*. Matrices in this ensemble are obtained by taking an m by m Fourier matrix and sampling n rows at random. Such matrices were studied by Candès et al. [13]; see also [19].

We note, however, that many more ensembles warrant exploration. Among these we count:

- *Random spikes and sinusoids*. Matrices in this ensemble are obtained by taking a concatenation of the m by m identity with the m by m Fourier matrix, and sampling n rows at random.
- *Random spikes and square waves*. Matrices in this ensemble are obtained by taking a concatenation of the m by m identity with the m by m Hadamard matrix, and sampling n rows at random.
- *Random orthogonal bases*. This ensemble is a generalization of the last two. Matrices in this ensemble are constructed by concatenating A random ortho-bases of dimensions n by n, so that the resulting matrix is n by An.

In the following, we find that the heuristic essentially provides a general scheme that can be applied to such ensembles, bounding the behavior of the LEBP.

5.1. Random signs ensemble

We studied the behavior of $v_0(A)$ for the random signs ensemble, as we did for the USE, and obtained

$$v_0(A) \approx 0.78 \cdot A^{-0.5},$$

as Fig. 5 illustrates. This is nearly identical to the fitted curve for the USE. Because our heuristic depends only on $v_0(A)$, it predicts the same breakdown points for the random signs ensemble as for the USE, i.e. the figures in Table 1 and the curves shown in Fig. 4 hold for the random signs ensemble as well.

To verify our predictions, Fig. 6 displays the percentage of perfect reconstructions versus k for

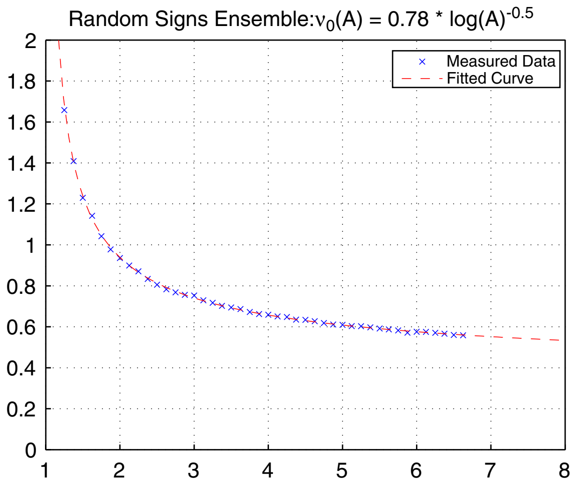

Fig. 5. Empirical estimate of the limit function $v_0(A)$ for the random signs ensemble.

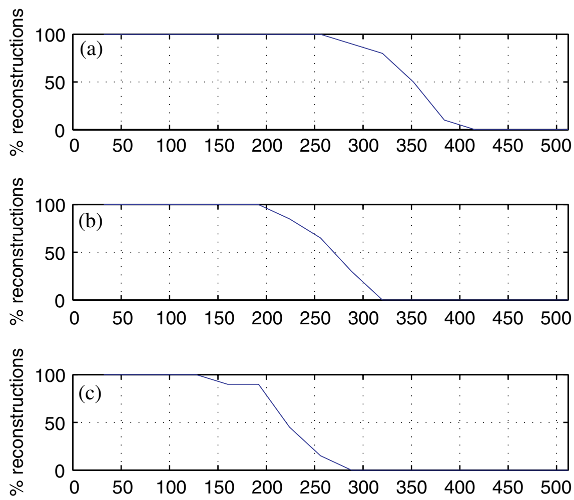

Fig. 6. Individual equivalence for the random signs ensemble. Panels (a)–(c) show percentage of perfect reconstructions versus the size of the support interval I for $n = 1024, m = 2048, 3072$ and 4096, respectively.

the random signs ensemble, summarizing the results of an experiment similar to the one described in Section 1.3 for the USE case. Panels (a)–(c) have plots for $n = 1024$ and $m = 2048, 3072$ and 4096, respectively. The empirical breakdown points are similar to those observed for the USE, and to our predictions.

5.2. Partial Hadamard and partial Fourier ensembles

We now consider the partial Hadamard and Fourier ensembles, which give results strikingly similar to each other, but also different from the other two ensembles.

We first ask if our heuristic even makes sense here. Recall that in deriving the heuristic, we implicitly assumed that the solution of (2.4) is in fact $V_{n,m-|I|} \cdot \|v_I\|_2$, where $V_{n,m}$ is the random variable representing the solution of $RP_1(n,m)$. This assumption can be justified reasonable for the USE and random signs ensemble, where different columns of Φ are stochastically independent. However, suppose we generate Φ by randomly sampling rows of an ortho-basis, like a Hadamard or Fourier basis. In that case, it is no longer trivial to assume that the solution of (2.4) behaves as $V_{n,m-|I|} \cdot \|v_I\|_2$. We now check this assumption.

We conducted the following experiment. We generate a dictionary Φ from the partial Fourier ensemble, and a random vector y uniform on S^{n-1}. We then solve

$$\min_{\delta} \|\delta\|_1 \quad \text{subject to} \quad \Phi\delta = y,$$

and denote the value of this problem $v_1(n,m)$. This is the familiar problem $RP_1(n,m)$, only with $\Phi_{n,m}$ drawn from a partial Fourier ensemble. In addition, given a random subset I with $|I| = k$, we pick y uniform on $\text{span}\{\Phi_I\} \cap \mathbf{S}^{n-1}$ and solve

$$\min_{\delta} \|\delta\|_1 \quad \text{subject to } \Phi_{I^c}\delta = y,$$

denoting the value $v_2(n,m;k)$; this is the problem appearing in (2.4), which we shall denote $(RP_2(n,m;k))$. For our heuristic to be valid on the partial Fourier ensemble, we need $v_1(n,m) \approx v_2(n,m-k)$.

Table 2 summarizes our findings. We see that v_1 and v_2 give very similar results, encouraging our use of the heuristic when Φ is drawn from the partial Fourier ensemble.

An additional finding is given in Fig. 7, which shows curves that were fitted to both sets of data, using the model $v(A) = C_1/\sqrt{\log(1 + C_2A)}$. The differences are indeed negligible.

Table 2
Comparison of solution values val($RP_1(n,m)$) and val($RP_2(n,m;k)$) for the partial Fourier ensemble

n	m_1	m_2	k	val(RP_1 (n,m))	val(RP_2 $(n,m;k)$)
512	1024	1054	30	15.03	15
512	1024	1084	60	15.03	14.85
512	1024	1114	90	15.03	14.7
512	1280	1310	30	14.32	14.28
512	1280	1340	60	14.32	14.16
512	1280	1370	90	14.32	14.03
512	1536	1566	30	13.64	13.73
512	1536	1596	60	13.64	13.57
512	1536	1626	90	13.64	13.58
512	1920	1950	30	13.02	12.96
512	1920	1980	60	13.02	13.04
512	1920	2010	90	13.02	12.96
512	2176	2206	30	12.79	12.7
512	2176	2236	60	12.79	12.69
512	2176	2266	90	12.79	12.67
512	2560	2590	30	12.35	12.26
512	2560	2620	60	12.35	12.31
512	2560	2650	90	12.35	12.25
512	3072	3102	30	11.9	11.89
512	3072	3132	60	11.9	11.84
512	3072	3162	90	11.9	11.86

Similar results are obtained for the partial Hadamard ensemble. Fig. 8 shows the curves fitted to $v_1(A)$ and $v_2(A)$. Test curves were generated in a manner similar to that described above. Again, we see that the two sets of results have numerically similar behavior. Further, we observe that the limit function for the partial Hadamard ensemble is nearly identical to the limit function for the partial Fourier ensemble.

While the results for the partial Fourier ensemble and the partial Hadamard ensemble are indeed similar, the behavior of $\hat{v}_0(A)$ is markedly different in these cases than the behavior seen for the uniform and the random signs ensembles. This raises the natural question: With this new behavior of $\hat{v}_0(A)$, will the heuristic accurately predict the LEBP?

Indeed, our experiments show that when using the obtained limit function, our heuristic reliably predicts $\rho_{L^+}(A)$ and $\rho_{L^-}(A)$ for these two ensembles. Table 3 summarizes our simulation results.

To gain insight about the behavior of the LEBP bounds for a wide range of A, we followed the guidelines of Section 4 and fitted the model $C_1/\log(1+C_2A)$ to the computed $\rho_{L^+}(A)$ and $\rho_{L^-}(A)$, yielding the approximations

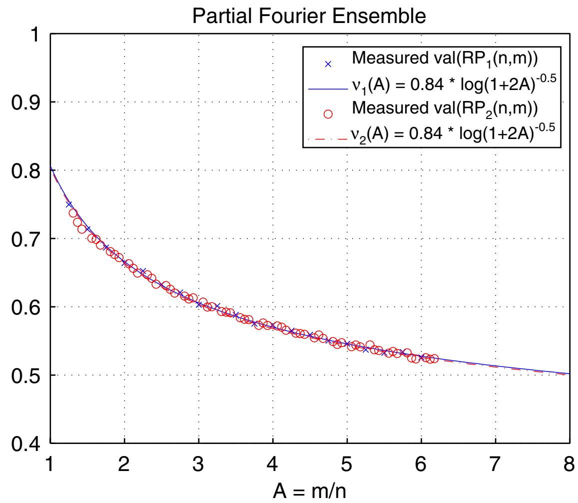

Fig. 7. Comparison of the behavior of val($RP_1(n,m)$) and val($RP_2(n,m;k)$) for the partial Fourier ensemble. The solid line shows the curve fitted to $v_1(A)$, the dash-dotted line shows the curve fitted to $v_2(A)$.

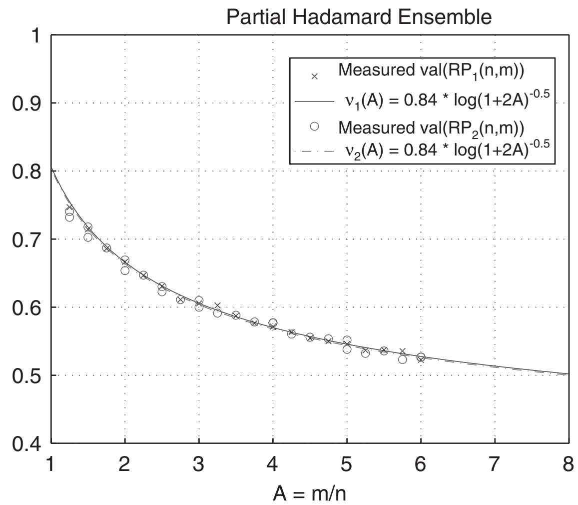

Fig. 8. Comparison of the behavior of $RP_1(n,m)$ and $RP_2(n,m;k)$ for the partial Hadamard ensemble. The solid line shows the curve fitted to $v_1(A)$, the dash-dotted line shows the curve fitted to $v_2(A)$.

$$\rho_{L^+}(A) = 0.65/\log(1+10A),$$

Table 3
Predicted versus empirical breakdown point for the partial Fourier and partial Hadamard ensembles. ρ_{L^+} and ρ_{L^-} were obtained by solving Eqs. (2.2) and (2.10), respectively

n	m	A	ρ_{L^+}	ρ_{L^-}	Empirical LEBP (Fourier)	Empirical LEBP (Hadamard)
512	1024	2	0.21	0.16	0.17	0.15
512	1536	3	0.18	0.14	0.13	0.12
512	2048	4	0.17	0.13	0.1	0.1
1024	2048	2	0.21	0.16	0.16	0.17
1024	3072	3	0.18	0.14	0.12	0.12
1024	4096	4	0.17	0.13	0.09	0.1

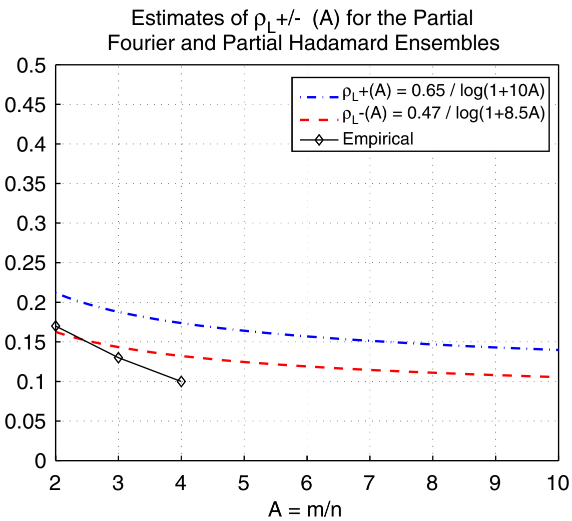

Fig. 9. Estimates of heuristic LEBP bounds for the partial Fourier and partial Hadamard ensembles. Solid curve shows empirical breakdown values, dashed curve shows ρ_{L^-}, dash-dotted curve shows ρ_{L^+}.

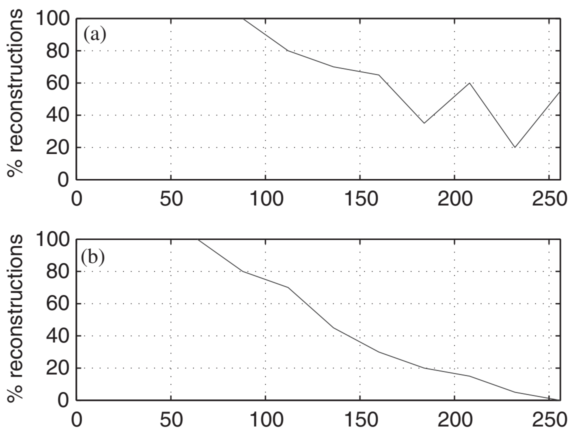

Fig. 10. Individual equivalence for the partial Fourier and partial Hadamard ensembles. Panels (a) and (b) show percentage of perfect reconstructions versus the size of the support $|I|$ in the case $n = 512, m = 1024$ for the partial Fourier and partial Hadamard ensembles, respectively.

$$\rho_{L^-}(A) = 0.47/\log(1 + 8.5A),$$

for the partial Fourier ensemble (since the estimates for the partial Hadamard ensemble are nearly identical to the estimates for the partial Fourier ensemble, these results hold for the partial Hadamard ensemble as well). The results are depicted in Fig. 9.

To better visualize the empirical breakdown point for these two matrix ensembles, Fig. 10, panels (a) and (b) show the percentage of perfect reconstructions versus the support size $|I|$ for the partial Fourier and partial Hadamard ensembles, in the case $n = 512, m = 1024$.

6. Conclusions

We studied the ensemble of n by m matrices Φ which have unit-norm columns in ℓ^2, where $n<m<An$. We considered the phenomenon of ℓ^1–ℓ^0 equivalence; namely, that, for certain sparse vectors α_0 having (say) k nonzeros, a 'typical' n by m matrix Φ, $m<An$, the data $y = \Phi\alpha_0$ afford

instances of problems (P_1) and (P_0) with the same solution, namely α_0. We identified three different notions of equivalence—global: (for all α_0 with at most k nonzeros); individual (most α_0 with at most k nonzeros); local (most support sets $supp(\alpha_0)$ and all vectors sharing that common support). The last two notions—*individual* and *local* equivalence—are important because they refer to computationally verifiable (hence empirically measurable) phenomena. We focus on the analysis of local equivalence and give heuristic formulas suggesting that (for large n) breakdown of local equivalence happens for arbitrarily supported objects with k nonzeros, in an interval defined by

$$\rho_{L^-}(m/n)(1-\varepsilon) \lesssim k/n < \rho_{L^+}(m/n) \cdot (1+\varepsilon),$$

with the lower bound being non-rigorously supported. These heuristic formulas were empirically tested by an extensive simulation, generating many random sparse vectors with different supports and different degrees of sparsity and checking whether the solution of (P_1) was indeed the preferred sparsest solution.

We see three possible applications of the heuristic: first, in generating a simple approximate formula for the breakdown point for matrix ensembles of current interest; second, in the predicting breakdown points of new matrix ensembles; and third, in suggesting new theoretical analyses.

In this paper we explored the application of the heuristic to several matrix ensembles beyond the USE. We applied the heuristic to the ensemble of n by m matrices with random ± 1 entries, with completely parallel results to the USE case. On the other hand, for n by m matrices drawn from the partial Hadamard and partial Fourier ensembles, the heuristic formulas predict different breakdown points than the USE. The predictions are systematically smaller but again accurate.

We also show that our heuristic formula can be converted into rigorous in-probability bounds simply by a rigorous proof of (2.9).

Acknowledgments

Partial support from NSF DMS 00-77261, and 01-40698 (FRG) and ONR-MURI. Thanks to Michael Saunders and Emmanuel Candès for helpful discussions.

Appendix

To understand why ρ_{L^-} cannot be rigorously supported let us try to derive it along the lines of the upper bound. Let α_0 be any vector with support I, where $|I|/n < \rho_{L^-}(m/n) - \varepsilon$. Consider all perturbations δ lying in the nullspace of Φ, i.e. $\Phi\delta = 0$. We will show that every such perturbation of α_0 increases the ℓ^1 norm. Hence α_0 is the global solution of the ℓ^1 problem.

For small $t > 0$,

$$\begin{aligned}\|\alpha_0 + t \cdot \delta\|_1 - \|\alpha_0\|_1 &= t \sum_{i \in I} \operatorname{sgn}(\alpha_0(i))\delta_I(i) \\ &\quad + t \sum_{i \in I^c} |\delta_I(i)| \\ &\geqslant t \cdot (\|\delta_{I^c}\|_1 - \|\delta_I\|_1).\end{aligned}$$

Hence, if

$$\|\delta_{I^c}\|_1 > \|\delta_I\|_1$$

for such a perturbation, it must increase the objective value. It is needed to show that this to hold for all such perturbations.

Note that for any perturbation,

$$\|\delta_I\|_1 \leqslant \sqrt{|I|}\|\delta_I\|_2,$$

and, for $v_I = \Phi_I \delta_I$

$$\|v_I\|_2 \geqslant \lambda_{\min}^{1/2} \|\delta_I\|_2.$$

As above, let $\rho_{L^-} = \rho_{L^-}(A)$ solve Eq. (2.10). Then, if $k < n(\rho_{L^-}(m/n) - \varepsilon)$ we have

$$\begin{aligned}\|\delta_I\|_1 &\leqslant \sqrt{|I|}\|\delta_I\|_2 = \sqrt{\rho n}\|\delta_I\|_2 \\ &< \sqrt{n} \cdot v_0(A-\rho) \cdot (1-\sqrt{\rho})\|\delta_I\|_2 \\ &\sim v(n, m-|I|)\|v_I\|_2 \lesssim \|\delta_{I^c}\|_1.\end{aligned}$$

Here everything is on sound footing until the symbol $\lesssim$. This is non-rigorous, because if we looked closer at the underlying arguments we would see that the needed quantity is not $v(n, m - |I|)$ but an alternate (and still smaller) quantity we will call $\tilde{v}(n, m-|I|; |I|)$.

To define this, let X_I denote a random uniform subspace of $\mathbf{R}^n$ of dimension $|I|$. Let

$S_I^{k-1} = X_I \cap S^{n-1}$. Then

$$(\tilde{RP}_1(n,m)) \quad \min \|\alpha\|_1 \quad \text{subject to } \Phi\alpha = x, \quad x \in S_I^{k-1}.$$

In short, we look for the smallest value of (RP_1) over all x's of unit length in a random subspace of dimension k. Define now $\tilde{v}(n, m-|I|;|I|) = \mathrm{med}(\mathrm{val}(\tilde{RP}_1(n,m)))$; clearly $\tilde{v}(n, m-|I|;|I|) \leqslant v(n, m-|I|)$. Supposing the existence of the conjectural limit

$$\tilde{v}_0(A,\rho) = \lim_{n\to\infty} \tilde{v}(n, \lfloor An - \rho n \rfloor; \lfloor \rho n \rfloor), \tag{A.1}$$

the formal solution $\tilde{\rho}_{L^-}$ to

$$\frac{\sqrt{\rho}}{(1-\sqrt{\rho})} = \tilde{v}_0(A;\rho) \tag{A.2}$$

can be put on a rigorous footing. Thus, one could show by the above arguments that, conditional on (A.1), for sufficiently small $\varepsilon > 0$, and sufficiently large n, for $k < n(\tilde{\rho} - \varepsilon)$ we have, with probability exceeding $1-\varepsilon$, the following display, uniform in δ_I:

$$\|\delta_I\|_1 \cdot (1+\varepsilon) \leqslant \tilde{v}(n, m-k;k)\|v_I\|_2 < \|\delta_{I^c}\|_1(1-\varepsilon).$$

This, as indicated above, would prove that any α_0 with support I must be the corresponding minimum ℓ^1 solution.

This more involved line of argument yields a rigorously motivated lower bound $\tilde{\rho}_{L^-}$. Unfortunately, $\tilde{v}(n, m-k;k)$ is not directly accessible to computation, although it is a geometrically very natural and interesting quantity, associated with Milman's quotient of a subspace (QS) theorem [26].

How much smaller is $\tilde{v}(n, m-k;k)$ than $v(n, m-k)$? An argument, based on the QS theorem, suggests that the value of the solution will be smaller by a factor $(1-\sqrt{k/n})\log(n/k)$ than the value of the solution of $(\tilde{RP}_1(n, m-k))$.

In practice, we use ρ_{L^-}, which is computationally accessible, trivial to compute in conjunction with the upper bound ρ_{L^+}, and roughly accurate.

References

[1] R.R. Coifman, Y. Meyer, S. Quake, M.V. Wickerhauser, Signal processing and compression with wavelet packets, in: J.S. Byrnes, J.L. Byrnes, K.A. Hargreaves, K. Berry (Eds.), Wavelets and Their Applications, 1994.

[2] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.

[3] S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1999) 33–61.

[4] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995) 227–234.

[5] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory 47 (7) (November 2001) 2845–2862.

[6] D.L. Donoho, M. Elad, Optimally sparse representation from overcomplete dictionaries via ℓ^1 norm minimization, Proc. Nat. Acad. Sci. USA 100 (5) (March 4, 2002) 2197–2202.

[7] M. Elad, A.M. Bruckstein, A generalized uncertainty principle and sparse representations in pairs of bases, IEEE Trans. Inform. Theory 49 (2002) 2558–2567.

[8] J.J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Inform. Theory 50 (6) (June 2004) 1341–1344.

[9] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (12) (December 2003) 1320–1325.

[10] J.A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (11) (October 2004) 2231–2242.

[11] J.A. Tropp, Just relax: convex programming methods for subset selection and sparse approximation, Manuscript.

[12] D.M. Malioutov, M. Çetin, A.S. Willsky, Optimal sparse representations in general overcomplete bases, IEEE International Conference on Acoustics, Speech and Signal Processing, May 2004, Montreal, Canada.

[13] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, Manuscript, 2004.

[14] D.L. Donoho, For most large underdetermined systems of linear equations, the minimal ℓ^1 solution is also the sparsest solution, Manuscript, submitted for publication, URL: http://www-stat.stanford.edu/~donoho/Reports/2004.

[15] T. Strohmer, R. Heath Jr., Grassmannian frames with applications to coding and communications, Appl. Comput. Harm. Anal. 14 (3) (2003) 257–275.

[16] E.J. Candès, T. Tao, Near optimal signal recovery from random projections: universal encoding strategies? Manuscript, 2004.

[17] D.L. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory, to appear.

[18] D.L. Donoho, Compressed sensing. Manuscript, submitted for publication, URL: http://www-stat.stanford.edu/~donoho/Reports/2004.

[19] Y. Tsaig, D.L. Donoho, Extensions of compressed sensing, Signal Process., to appear.

[20] P. Billingsley, Probability and Measure, Wiley, New York, 1995.

[21] D.L. Donoho, For most underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution, Manuscript, submitted for publication, URL: http://www-stat.stanford.edu/~donoho/Reports/2004.

[22] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. 9 (1988) 543–560.

[23] B.S. Kashin, Diameters of certain finite-dimensional sets in classes of smooth functions, Izv. Akad. Nauk SSSR, Ser. Mat. 41 (2) (1977) 334–351.

[24] A.Y. Garnaev, E.D. Gluskin, On widths of the Euclidean ball, Soviet Math. Dokl. 30 (1984) 200–203 (in English).

[25] M. Harwit, N.J.A. Sloane, Hadamard Transform Optics, Academic Press, New York, 1979.

[26] G. Pisier, The Volume of Convex Bodies and Banach Space Geometry, Cambridge University Press, Cambridge, 1989.