Restricted Isometry Constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$

Michael E. Davies Rémi Gribonval

Abstract

This paper investigates conditions under which the solution of an underdetermined linear system with minimal ℓ^p norm, $0 < p \leq 1$, is guaranteed to be also the sparsest one. The results highlight the pessimistic nature of sparse recovery analysis when recovery is predicted based on the restricted isometry constants (RIC) of the associated matrix. Matrices are constructed with RIC δ_{2m} arbitrarily close to $1/\sqrt{2} \approx 0.717$ where sparse recovery with $p = 1$ fails for at least one m-sparse vector. This indicates that there is limited room for improving over the best known positive results of Foucart and Lai, which guarantee that ℓ^1-minimisation recovers all m-sparse vectors for any matrix with $\delta_{2m} < 2(3 - \sqrt{2})/7 \approx 0.4531$. Another consequence of the construction is that recovery conditions expressed uniformly for all matrices in terms of RIC must require that all $2m$-column submatrices are extremely well conditioned (condition numbers less than 2.5). In contrast, matrices are also constructed with δ_{2m} arbitrarily close to one and $\delta_{2m+1} = 1$ where ℓ^1 minimisation succeeds for any m-sparse vector. This illustrates the limits of RIC as a tool to predict the behaviour of ℓ^1 minimisation. These constructions are a by-product of tight conditions for ℓ^p recovery ($0 \leq p \leq 1$) with matrices of unit spectral norm, which are expressed in terms of the minimal singular values of $2m$-column submatrices. The results show that, compared to ℓ^1 minimisation, ℓ^p minimisation recovery failure is only slightly delayed in terms of the RIC values. Furthermore in this case the minimisation is nonconvex and it is important to consider the specific minimisation algorithm being used. It is shown that when ℓ^p optimisation is attempted using an iterative reweighted ℓ^1 scheme, failure can still occur for δ_{2m} arbitrarily close to $1/\sqrt{2}$.

Index Terms

underdetermined linear system, sparse representation, overcomplete dictionary, compressed sensing, inverse problem, restricted isometry property, convex optimisation, nonconvex optimisation, iterative reweighted optimisation.

I. Introduction and state of the art

This paper investigates conditions under which the solution $\hat{\mathbf{y}}$ of minimal ℓ^p norm, $0 < p \leq 1$, of an underdetermined linear system $\mathbf{x} = \mathbf{\Phi}\mathbf{y}$ is guaranteed to be also the sparsest one. This is a central problem in sparse overcomplete signal representations, where $\mathbf{x}$ is a vector representing some signal or image, $\mathbf{\Phi}$ is an overcomplete

IDCOM & Joint Research Institute for Signal and Image Processing Edinburgh University, King's Buildings, Mayfield Road, Edinburgh EH9 3JL, Tel. +44 (0)131 650 5795, email: mike.davies@ed.ac.uk

METISS project-team, INRIA Rennes, Campus de Beaulieu, F-35042 Rennes Cedex, France, Tel. +33 (0)299 84 25 06, email: remi.gribonval@irisa.fr

signal dictionary, and $\mathbf{y}$ is a sparse representation of the signal. This problem is also at the core of compressed sensing, where $\mathbf{\Phi}$ is called a sensing matrix, $\mathbf{x}$ is a collection of M linear measurements of some ideally sparse data $\mathbf{y}$. Although in both settings it is practically relevant to consider sparse *approximation* rather than exact sparse *representation*, most of the results of this paper are of a negative nature and naturally extend from the representation setting chosen here (for the sake of simplicity) to the approximation setting.

The proposed approach is twofold:

- we construct matrices (which we will call *dictionaries* from now on) $\mathbf{\Phi}$ with "good" restricted isometry properties where sparse recovery with ℓ^p minimisation will nevertheless fail for at least one sparse vector.
- we construct dictionaries $\mathbf{\Phi}$ with "bad" restricted isometry properties where sparse recovery with ℓ^p minimisation will nevertheless succeed for all (sufficiently) sparse vectors.

The goal is to understand how much improvement is possible over the best known positive results which relate restricted isometry constants to sparse ℓ^p recovery.

A. Notations

Given a vector $\mathbf{x} \in \mathbb{R}^M$ and a matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M < N$, we are interested in sparse solutions to

$$\mathbf{x} = \mathbf{\Phi}\mathbf{y} \tag{1}$$

We will denote by $\|\mathbf{y}\|_p$ the ℓ^p sparsity measure defined as:

$$\|\mathbf{y}\|_p := \big(\sum_{j=1}^{N} |y_j|^p\big)^{1/p} \tag{2}$$

where $0 < p \leq 1$. When $p = 0$, $\|\mathbf{y}\|_0$ denotes the ℓ^0 pseudo-norm that counts the number of non-zero elements of $\mathbf{y}$. The coefficient vector $\mathbf{y}$ is said to be m-sparse if $\|\mathbf{y}\|_0 \leq m$.

We will use $\mathcal{N}(\mathbf{\Phi})$ for the null space of $\mathbf{\Phi}$. We will also make use of the subscript notation $\mathbf{y}_\Omega$ to denote a vector that is equal to some $\mathbf{y}$ on the index set Ω and zero everywhere else. Denoting $|\Omega|$ the cardinality of Ω, the vector $\mathbf{y}_\Omega$ is $|\Omega|$-sparse and we will say that the support of the vector $\mathbf{y}$ lies within Ω whenever $\mathbf{y}_\Omega = \mathbf{y}$. For matrices the subscript notation $\mathbf{\Phi}_\Omega$ will denote a submatrix composed of the columns of $\mathbf{\Phi}$ that are indexed in the set Ω.

B. Known conditions for ℓ^p sparse recovery

It has been shown in [14] that if:

$$\|\mathbf{z}_\Omega\|_p < \|\mathbf{z}_{\Omega^c}\|_p \tag{3}$$

holds for all nonzero $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ then any vector $\mathbf{y}^\star$ whose support lies within Ω, can be recovered as *the unique* solution of the following optimisation problem (which is non-convex for $0 \leq p < 1$):

$$\hat{\mathbf{y}} = \operatorname*{argmin}_{\mathbf{y}} \|\mathbf{y}\|_p \text{ s.t. } \mathbf{\Phi}\mathbf{y} = \mathbf{\Phi}\mathbf{y}^\star. \tag{4}$$

Furthermore this condition, which is often referred to as the "null space property", is tight in the sense that if the inequality (3) does not hold for some $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ then the vector $\mathbf{y}^\star := \mathbf{z}_\Omega$ is supported on Ω but is not the unique minimiser of (4), a property that we will refer to as ℓ^p failure. As a consequence, if (3) holds for all $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ and all index sets Ω of size m, then any m-sparse vector $\mathbf{y}^\star$ is recovered as the unique minimiser of (4). This condition is again tight, and it has been shown in [15], [16] that when it is satisfied for some $0 < p \leq 1$ it is also satisfied for all $0 \leq q \leq p$.

Using (4), particularly when $p = 1$, has become a popular mean of solving for sparse representations. This is partly due to empirical evidence [5] that it often performs well and partly due to theoretical results [2], [3], [6], [14], [17]. An important concept in this regard that has been particularly influential in the emerging field of compressed sensing is the restricted isometry constant (RIC), δ_k. For a matrix $\mathbf{\Phi}$ this is defined as the smallest number such that:

$$(1 - \delta_k) \leq \frac{\|\mathbf{\Phi} \mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \leq (1 + \delta_k) \tag{5}$$

for every vector $\mathbf{y}$ and every index set Ω with $|\Omega| \leq k$. One weakness of the RIC is that the upper bound and the lower bound play fundamentally different roles and it is not preserved under a re-scaling of the dictionary [11] while recovery properties clearly are. One can, however, usually overcome the latter problem by considering an appropriately re-scaled dictionary such that both the upper and the lower bound are tight.

The RIC's importance can be linked with the following results:

1) Every m-sparse representation is unique if and only if [7]

$$\delta_{2m} < 1 \tag{6}$$

for an appropriately re-scaled dictionary. Furthermore almost every dictionary $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M \geq 2m$ satisfies this condition (again with appropriate re-scaling). Foucart and Lai [11] have also shown that for a given dictionary with $\delta_{2m+2} < 1$ there exists a sufficiently small p for which solving (4) is guaranteed to recover any m-sparse vector.

2) If

$$\delta_{2m} < 2(3 - \sqrt{2})/7 \approx 0.4531 \tag{7}$$

then every m-sparse representation can be exactly recovered using linear programming to solve (4) with $p = 1$, [11]. Furthermore most dictionaries $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ (sampled from an appropriate probability model) will have an RIC $\delta_{2m} < \delta$ as long as: $M \geq C\delta^{-1} m \log(N/m)$, where C is some constant [1].

The RIC also bounds the condition number, κ, of submatrices, $\mathbf{\Phi}_\Omega$, of a dictionary,

$$\kappa(\mathbf{\Phi}_\Omega) \leq \sqrt{\frac{1 + \delta_k}{1 - \delta_k}}, \qquad |\Omega| \leq k \tag{8}$$

(indeed, Foucart and Lai [11] formulated their results in asymmetric bounds $\alpha_k \leq \|\mathbf{\Phi} \mathbf{y}_\Omega\|_2^2 / \|\mathbf{y}_\Omega\|_2^2 \leq \beta_k$ that provide a sharper bound on the maximal submatrix condition number, with less re-scaling issues). This in turn bounds the Lipschitz constant of the inverse mapping resulting from solving the optimisation problem (4). In this regard the RIC also plays an important role in the noisy recovery problems [3], [11]: $\mathbf{x} = \mathbf{\Phi y} + \epsilon$ or $\mathbf{x} = \mathbf{\Phi}(\mathbf{y} + \epsilon)$ where ϵ is an unknown but bounded noise term.

Note that when (7) holds all the $2m$-submatrices have condition number $\kappa(\mathbf{\Phi}_\Omega) \leq 1.7$ when $|\Omega| \leq 2m$, so they are extremely well behaved. In contrast, $\delta_{2m} < 1$ imposes the finiteness of the condition number of the submatrices as only constraint.

C. Contributions

The bound (7) is an improvement over previous known bounds for ℓ^1 recovery [3]. However, in the proof of these bounds [3], [11], there are a number of estimates that are not tight. It is therefore an open question as to how much better we could expect to do, i.e. how large can we set $\delta \leq 1$ while still guaranteeing ℓ^1 recovery of any m-sparse vector for any dictionary with $\delta_{2m} < \delta$? This question is partially addressed by the following result:

Theorem 1: For any $\epsilon > 0$ there exists an integer m and a dictionary $\mathbf{\Phi}$ with a restricted isometry constant $\delta_{2m} \leq 1/\sqrt{2} + \epsilon$ for which ℓ^1 recovery fails on some m-sparse vector.

The main idea of the proof is to first reduce the search for a failing dictionary to so-called *minimally redundant* (i.e., $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M = N - 1$) unit spectral norm dictionaries. Such dictionaries have one-dimensional kernels which simplifies the calculation considerably. They are however of little practical interest for Compressed Sensing, since the number of measurements is about the same as the dimension of the signals. The proof is by an explicit construction which we will develop in the next section and is a by-product of some more general result concerning certain isometry conditions for which ℓ^p recovery fails, $0 < p \leq 1$. Indeed our complete results for RIC recovery conditions along with the result of [11] are summarised graphically in Figure 1.

The plot is divided up into three regions. Dictionaries in the bottom region [11] are guaranteed to succeed using any ℓ^p optimisation. In the top region there exist dictionaries, specifically minimally redundant row orthonormal dictionaries (also sometimes called unit norm tight frames) that are guaranteed to fail to recover at least one m-sparse vector $\mathbf{y}$ (Theorem 3). On the other hand, we can also find dictionaries (again minimally redundant row orthonormal dictionaries) that are ℓ^p succeeding for any $0 < p \leq 1$ with a RIC, δ_{2m}, arbitrarily close to one (Lemma 1).

Although there is a gap between the positive result of Foucart and Lai [11] for $p = 1$ and the negative result presented here, it is not a large one. For example, even if the positive result could be tightened to $\delta_{2m} < 1/\sqrt{2}$ (which would be the case if our negative results happened to be sharp, and the result *is* sharp for $2m > N - M$ with mildy overcomplete row orthonormal dictionaries, see Corollary 1 below) this would still require that the condition numbers of any $2m$-column submatrix of $\mathbf{\Phi}$ would have to be $\kappa(\mathbf{\Phi}_\Omega) \leq 2.5$, for $|\Omega| \leq 2m$, which from any perspective is still extremely well conditioned.

The plot suggests that there might be some benefit in using $p \ll 1$ to improve sparse recovery. However in this case the optimisation problem is no longer convex and so we need to consider *algorithm specific* recovery results. In this paper we examine the iterative reweighted ℓ^1 technique proposed in [9], [10], [4], [11] and present the following complement to Theorem 1.

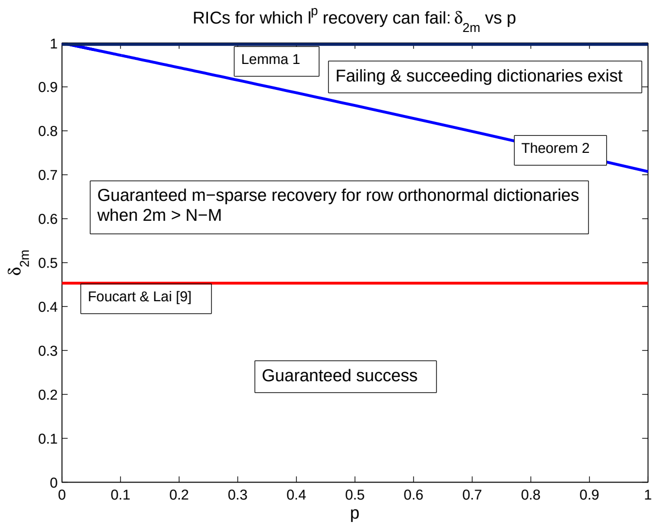

Fig. 1. A summary of known results ([11], Lemma 1 and Theorem 3) relating the restricted isometry constant to ℓ^p recovery. Note the guaranteed success for row orthonormal dictionaries is only applicable to mildly overcomplete dictionaries since by definition $N < 2M$ (see Corollary 1).

Theorem 2: For any $\epsilon > 0$ there exists an integer m and a dictionary $\mathbf{\Phi}$ with a restricted isometry constant $\delta_{2m} \leq 1/\sqrt{2} + \epsilon$ for which recovery using any iteratively reweighted ℓ^1 algorithm fails on some m-sparse vector.

This is a somewhat surprising result since one would suspect that the adaptivity built into the choice of the weight should enable improved performance. This result does not necessarily imply however that the uniform performance of iterative reweighted ℓ^1 techniques is no better than ℓ^1 minimisation (although we suspect that the empirically observed benefits of such algorithms are more likely to be due to the presence of a range of coefficient scales). Instead the result highlights the danger of characterising sparse recovery uniformly in terms of the RIP.

The rest of the paper is structured as follows. In section II we introduce a variation on the classical RIC. We then develop our RIC results based upon an explicit minimally redundant unit spectral norm dictionary construction. In section IV we explore our results numerically for both high dimensional dictionaries and a simple 1-sparse low dimensional example. Finally we examine the class of ℓ^p optimisation algorithms based upon iterative reweighted ℓ^1. We conclude the paper with a discussion of implications of these results.

II. Isometry measures for unit spectral norm dictionaries

We will find it convenient to work with a slightly stronger condition than the usual restricted isometry property (RIP), one associated with **unit spectral norm dictionaries**, i.e. dictionaries such that

$$|||\mathbf{\Phi}||| := \sup_{\mathbf{y} \neq 0} \frac{\|\mathbf{\Phi y}\|_2}{\|\mathbf{y}\|_2} = 1. \tag{9}$$

Definition 1 (asymmetric RIC): Given a unit spectral norm dictionary $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ let the asymmetric RIC σ_k^2

be defined as:

$$\sigma_k^2(\mathbf{\Phi}) := \min_{\substack{\mathbf{y}_\Omega \\ |\Omega| \leq k}} \frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \tag{10}$$

We will usually drop the dependence on $\mathbf{\Phi}$ when it is unambiguous. Clearly, as the maximum of any submatrix squared singular value is bounded by 1:

$$\max_{\substack{\mathbf{y}_\Omega \\ |\Omega| \leq k}} \frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \leq 1, \tag{11}$$

a unit spectral norm dictionary $\mathbf{\Phi}$ with a given σ_k^2 implies the existence of a re-scaled dictionary, $\mathbf{\Psi}_k$:

$$\mathbf{\Psi}_k := \left(\frac{2}{1+\sigma_k^2}\right)^{\frac{1}{2}} \mathbf{\Phi} \tag{12}$$

with a RIC, δ_k:

$$\delta_k(\mathbf{\Psi}_k) \leq \frac{1-\sigma_k^2(\mathbf{\Phi})}{1+\sigma_k^2(\mathbf{\Phi})} \tag{13}$$

Equality does not always hold for the re-scaled dictionary, since equality in (13) requires equality in (11). Under certain circumstances, however, equality can be assured.

Remark 1 (Condition for equality in (13) with row orthonormal dictionaries): Suppose that $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M \leq N$ is a *row orthonormal dictionary*. Then for any $k > N - M$ we have:

$$\delta_k(\mathbf{\Psi}_k) = \frac{1-\sigma_k^2(\mathbf{\Phi})}{1+\sigma_k^2(\mathbf{\Phi})} \tag{14}$$

where $\mathbf{\Psi}_k$ is defined in (12). Moreover, $\mathbf{\Psi}_k$ is the optimal re-scaling of $\mathbf{\Phi}$ with respect to the RIC δ_k in the sense that $\delta_k(\alpha\mathbf{\Phi}) \geq \delta_k(\mathbf{\Psi}_k)$ for any $\alpha > 0$. In particular for minimally redundant row orthonormal dictionaries (i.e., when $M = N - 1$) equality (14) is true for any $k \geq 2$.

Proof:

$\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{Id}$. For every vector $\mathbf{y} \in \mathbb{R}^N$, defining $\mathbf{x} := \mathbf{\Phi}\mathbf{y}$ and $\mathbf{z} := \mathbf{y} - \mathbf{\Phi}^T\mathbf{\Phi}\mathbf{y}$ yields an orthogonal decomposition $\mathbf{y} = \mathbf{\Phi}^T\mathbf{x} + \mathbf{z}$ hence

$$\|\mathbf{\Phi}\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 = \|\mathbf{\Phi}^T\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2 - \|\mathbf{z}\|_2^2$$

and the upper bound in (11) is therefore achieved as long as we can find a $\mathbf{y}_\Omega$ that is in the range of $\mathbf{\Phi}^T$. For any Ω of size k, the dimension of the subspace spanned by all vectors of the form $\mathbf{y}_\Omega$ is k while the codimension of the range of $\mathbf{\Phi}^T$ is $N - M$. Hence if $k > N - M$ there exists at least one nonzero vector in the intersection of these subspaces. The optimality of the re-scaled dictionary $\mathbf{\Psi}_k$ follows from the tightness of both upper and lower bounds in (5) for $\mathbf{\Psi}_k$. ■

III. Dictionaries with small δ_{2m} where ℓ^p can fail

Our aim is to construct dictionaries $\mathbf{\Phi}$ where sparse recovery will fail for at least one m-sparse vector $\mathbf{y} \in \mathbb{R}^N$. We consider the ℓ^p problem for any $0 < p \leq 1$ although we only provide closed form results for ℓ^1. We are therefore looking for dictionaries that explicitly fail the ℓ^p recovery condition (3) while possessing small RIC δ_{2m}.

To find 'ℓ^p failing dictionaries' (i.e., dictionaries for which ℓ^p minimisation fails to recover at least one m-sparse vector[1]) with small RIC δ_{2m}, we will be looking for ℓ^p failing dictionaries with largest possible σ^2_{2m}. We will indeed prove somewhat more than Theorem 1, including tight results for ℓ^p failure with unit spectral norm dictionaries in terms of the asymmetric RIC σ^2_{2m}, and tight results for ℓ^p failure with row orthonormal dictionaries in terms of the RIC δ_{2m}.

Theorem 3: Consider $0 < p \leq 1$ and let $0 < \eta_p < 1$ be the unique positive solution to

$$\eta_p^{2/p} + 1 = \frac{2}{p}(1 - \eta_p) \tag{15}$$

- If $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a *unit spectral norm* dictionary and $2m \leq M < N$ and

$$\sigma^2_{2m}(\mathbf{\Phi}) > 1 - \frac{2}{2-p}\eta_p \tag{16}$$

  then all m-sparse vectors can be uniquely recovered by solving (4).
- For every $\epsilon > 0$, there exist integers $m \geq 1, N \geq 2m+1$ and a minimally redundant row orthonormal dictionary $\mathbf{\Phi} \in \mathbb{R}^{(N-1)\times N}$ with:

$$\sigma^2_{2m}(\mathbf{\Phi}) \geq 1 - \frac{2}{2-p}\eta_p - \epsilon \tag{17}$$

  for which there exists an m-sparse vector which cannot be uniquely recovered by solving (4).

Whenever η_p is irrational the inequality in (16) can be replaced with $\geq$. Whenever η_p is rational, equality can be achieved with $\epsilon = 0$ in (17).

Specialising to $p = 1$ we have $\eta_1^2 + 2\eta_1 - 1 = 0$, hence $\eta_1 = \sqrt{2} - 1$ and the right-hand side in (16) is $3 - 2\sqrt{2}$. In terms of the standard RIP δ_{2m} for the re-scaled dictionary (12) with $k = 2m$ this means, using (13), that for any $\epsilon > 0$ there exists a dictionary $\mathbf{\Psi}$ with $\delta_{2m} < 1/\sqrt{2} + \epsilon$ where ℓ^1 recovery can fail, and Theorem 1 is proved.

Combining Theorem 3 with Remark 1 above we get the following corollary:

Corollary 1: Assume that $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a suitably re-scaled *row orthonormal dictionary*. If

$$N - M < 2m \leq M < N \tag{18}$$

and

$$\delta_{2m}(\mathbf{\Phi}) < \frac{\eta_p}{2 - p - \eta_p} \tag{19}$$

then all m-sparse vectors can be uniquely recovered by solving (4). Whenever η_p is irrational, the inequality in (19) can be replaced with $\leq$.

Strictly speaking the condition $2m \leq M$ is redundant with (19) since $2m > M$ implies $\delta_{2m} \geq 1$.

By the second part of Theorem 3, Corollary 1 is sharp in the sense that the right-hand side in (19) cannot be weakened. This does not mean however that (19) is a necessary condition on the RIC for ℓ^p success, and there exist

[1]We will often omit the dependence on m when referring to 'ℓ^p failing dictionaries'.

dictionaries with δ_{2m} arbitratrily close to one which recover every m-sparse vector, as expressed by the following lemma.

Lemma 1: For any $\epsilon > 0$, there exist integers m and N and a minimally redundant row orthonormal dictionary $\mathbf{\Phi}_1 \in \mathbb{R}^{(N-1)\times N}$ along with re-scaled versions of $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$, such that every m-sparse vector is recovered by solving (4) with any of $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3$ and any $0 \leq p \leq 1$, yet

$$\sigma_{2m}^2(\mathbf{\Phi}_1) \leq \epsilon \tag{20}$$

$$\sigma_{2m+1}^2(\mathbf{\Phi}_1) = 0 \tag{21}$$

$$\delta_{2m}(\mathbf{\Phi}_2) > 1-\epsilon \tag{22}$$

$$\delta_{2m+1}(\mathbf{\Phi}_3) = 1. \tag{23}$$

Theorem 3 will be proved by explicitly constructing the ℓ^p failing unit spectral norm dictionaries with largest σ_{2m}^2 for a given pair (m, N) with $2m < N$, and a similar construction will be used to prove Lemma 1. We postpone the proofs and begin with a series of lemmatas.

Proposition 1 (Minimally redundant row orthonormal dictionaries are optimal among unit spectral norm dictionaries): Let $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ be an arbitrary unit spectral norm dictionary which is ℓ^p failing for some m-sparse vector with $M < N$. Then there exists a minimally redundant row orthonormal (unit spectral norm) dictionary $\mathbf{\Phi}^\star \in \mathbb{R}^{(N-1)\times N}$ which is ℓ^p failing for the same m-sparse vector such that for every k

$$\sigma_k^2(\mathbf{\Phi}) \leq \sigma_k^2(\mathbf{\Phi}^\star). \tag{24}$$

Proof: We consider the singular value decomposition: $\mathbf{\Phi} = V\Sigma U^T$ where $V \in \mathbb{R}^{M\times M}$ and $U^T \in \mathbb{R}^{M\times N}$ are row orthonormal, and $\Sigma \in \mathbb{R}^{M\times M}$ is diagonal. Since $\mathbf{\Phi}$ has unit spectral norm $|||\Sigma||| = 1$ and we have for any $\mathbf{y}$ and any Ω

$$\frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \leq \frac{\|U^T\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2}.$$

Since $\mathbf{\Phi}$ is ℓ^p failing for some m-sparse vector, by the characterisation (3), there exists some offending $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ and an index set Ω_m of size m such that

$$\|\mathbf{z}_{\Omega_m}\|_p^p \geq \|\mathbf{z}_{\Omega_m^c}\|_p^p \tag{25}$$

Now let $W \in \mathbb{R}^{N\times(N-M-1)}$ be an orthonormal basis over the orthogonal complement to $\{\mathbf{z}, U\}$, such that $\{\mathbf{z}, U, W\}$ forms an orthonormal basis over $\mathbb{R}^N$ ($\mathbf{z}$ is orthonormal to U since $\mathbf{\Phi}\mathbf{z} = 0$). We can then write any $\mathbf{y}_\Omega \in \mathbb{R}^N$ as:

$$\mathbf{y}_\Omega = \mathbf{z}a + U\mathbf{b} + W\mathbf{c}$$

for some $a \in \mathbb{R}, \mathbf{b} \in \mathbb{R}^M$ and $\mathbf{c} \in \mathbb{R}^{N-M-1}$. Define the minimally redundant row orthonormal dictionary $\mathbf{\Phi}^\star := [U, W]^T \in \mathbb{R}^{(N-1)\times N}$, which satisfies $|||\mathbf{\Phi}^\star||| = 1$. First, for any $\mathbf{y}_\Omega$ we have:

$$\frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \leq \frac{\|U^T\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} = \frac{\|\mathbf{b}\|_2^2}{a^2 + \|\mathbf{b}\|_2^2 + \|\mathbf{c}\|_2^2} \leq \frac{\|\mathbf{b}\|_2^2 + \|\mathbf{c}\|_2^2}{a^2 + \|\mathbf{b}\|_2^2 + \|\mathbf{c}\|_2^2} = \frac{\|\mathbf{\Phi}^\star\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2}$$

therefore $\sigma_k^2(\mathbf{\Phi}) \leq \sigma_k^2(\mathbf{\Phi}^\star)$. To conclude the proof we observe that by construction $\mathbf{\Phi}^\star\mathbf{z} = 0$, hence $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi}^\star)$, which combined with (25) and the characterisation (3) shows that $\mathbf{\Phi}^\star$ is ℓ^p failing for at least one m-sparse vector. ■

The proposition above shows that ℓ^p failing unit spectral norm dictionaries with largest σ^2_{2m} can be searched within the restricted set of ℓ^p failing minimally redundant row orthonormal dictionaries. We next evaluate the minimal singular values of the submatrices made of k columns of such $\mathbf{\Psi}$.

Proposition 2 (Minimal singular values of submatrices are characterised by the null space): Let $\mathbf{\Phi} \in \mathbb{R}^{(N-1)\times N}$ be a minimally redundant row orthonormal dictionary, and let $\mathbf{z} \in \mathbb{R}^N$ with $\|\mathbf{z}\|_2 = 1$ be a vector which spans $\mathcal{N}(\mathbf{\Phi})$. Denoting Ω_k the set indexing the k largest components of $\mathbf{z}$ we have for every k

$$\sigma^2_k(\mathbf{\Phi}) = 1 - \|\mathbf{z}_{\Omega_k}\|_2^2. \tag{26}$$

Proof: Since $\mathbf{\Phi}\mathbf{z} = 0$ and $\mathbf{\Phi}$ is row orthonormal, $\{\mathbf{z}, \mathbf{\Phi}^T\}$ forms an orthonormal basis in $\mathbb{R}^N$, and we can again write any vector $\mathbf{y}$ as:

$$\mathbf{y} = \mathbf{z}a + \mathbf{\Phi}^T\mathbf{b}$$

where $a \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^{N-1}$, and therefore $\|\mathbf{\Phi}\mathbf{y}\|_2^2 = \|\mathbf{b}\|_2^2$. If $\mathbf{y}$ has unit norm then

$$1 = \|\mathbf{y}\|_2^2 = a^2 + \|\mathbf{b}\|_2^2 = |\langle \mathbf{z}, \mathbf{y}\rangle|^2 + \|\mathbf{\Phi}\mathbf{y}\|_2^2$$

To find the minimal singular value associated with the submatrix $\mathbf{\Phi}_\Omega$ we need to solve the problem

$$\begin{aligned} \sigma^2_k(\mathbf{\Phi}) &= \min_{\Omega, |\Omega| \le k} \min_{\substack{\mathbf{y}_\Omega \\ \|\mathbf{y}_\Omega\|=1}} \|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2 \\ &= 1 - \max_{\Omega, |\Omega| \le k} \max_{\substack{\mathbf{y}_\Omega \\ \|\mathbf{y}_\Omega\|=1}} |\langle \mathbf{z}, \mathbf{y}_\Omega\rangle|^2 \end{aligned}$$

i.e., we need to find the unit vector $\mathbf{y}^\star_\Omega$ that is maximally correlated with $\mathbf{z}$. For a given Ω this is satisfied with $\mathbf{y}^\star_\Omega = \mathbf{z}_\Omega/\|\mathbf{z}_\Omega\|$, in which case $|\langle \mathbf{z}, \mathbf{y}_\Omega\rangle|^2 = \|\mathbf{z}_\Omega\|_2^2$. The best Ω is the one which captures the k largest components of $\mathbf{z}$, that is to say $\Omega^\star = \Omega_k$. ■

The proposition above shows that for minimally redundant row orthonormal dictionaries, $\sigma^2_k(\mathbf{\Phi})$ is completely determined by the unit vector $\mathbf{z}$ which spans the null space $\mathcal{N}(\mathbf{\Phi})$. Our original problem was to select an ℓ^p failing minimally redundant row orthonormal dictionary $\mathbf{\Phi}$ with maximal $\sigma^2_k(\mathbf{\Phi})$ for $k = 2m$. This is now turned into an optimisation problem where we wish to select a unit norm vector $\mathbf{z}$ that allows ℓ^p reconstruction failure for m-sparse vectors, while maximising σ^2_k, i.e. minimising $\|\mathbf{z}_{\Omega_k}\|_2^2$.

Before we proceed to characterize the structure of $\mathbf{z}$ further we need to introduce the following technical lemma that we will need below.

Lemma 2: Let $0 < p < 2$ and $u_1 > v_1 \ge v_2 > u_2 \ge 0$ such that $u_1^p + u_2^p = v_1^p + v_2^p$. Then $u_1^2 + u_2^2 > v_1^2 + v_2^2$.

Proof: Let $J = u_1^2 + u_2^2$ and $u_1^p + u_2^p = c$ for some constant $c > 0$. It is sufficient to show that $\partial J/\partial u_1 > 0$ whenever $u_1 > u_2$.

$$\frac{\partial J}{\partial u_1} = 2u_1 + 2u_2\frac{\partial u_2}{\partial u_1} = 2u_1 - 2u_2\left(\frac{u_2}{u_1}\right)^{1-p} = 2u_1\left(1 - \left(\frac{u_2}{u_1}\right)^{2-p}\right)$$

which is strictly positive if $u_1 > u_2 \ge 0$ and $p < 2$. ■

We next consider the precise form of $\mathbf{z}$ for failing dictionaries with maximal σ^2_{2k}.

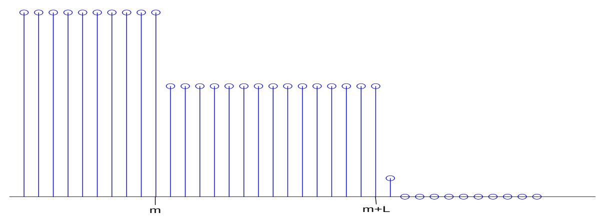

Fig. 2. A stylised depiction of an optimal vector from the null space for (27-30).

Without loss of generality, up to column permutation of $\mathbf{\Phi}$ and sign changes, we may assume that $z_i \geq z_{i+1} \geq 0$, and the ℓ^p failing assumption is that $\|\mathbf{z}_{\Omega_m}\|_p^p \geq \|\mathbf{z}_{\Omega_m^c}\|_p^p$. Defining $\Lambda_0 = \{1,\ldots,m\}$, $\Lambda_1 = \{m+1,\ldots,k\}$ and $\Lambda_2 = \{k+1,\ldots,N\}$, and with a little manipulation, the optimisation problem for finding a failing $\mathbf{z}$ with maximal associated σ_k^2 can be written in the form of (27-30) below. In particular, since (29) imposes the constraint $\|\mathbf{z}_{\Lambda_2}\|_2^2 = 1 - \|\mathbf{z}_{\Lambda_0}\|_2^2 - \|\mathbf{z}_{\Lambda_1}\|_2^2$, minimising $J(\mathbf{z})$ in (27) amounts to minimising $\|\mathbf{z}_{\Lambda_0}\|_2^2 + \|\mathbf{z}_{\Lambda_1}\|_2^2 = \|\mathbf{z}_{\Omega_k}\|_2^2$. The next lemma identifies the particularly simple form of the optimal vectors from the null space which is also depicted in Figure 2.

Lemma 3 (Shape of the optimal vector $\mathbf{z}$ *of the null space):* Consider $k \geq 2m$ and let $\mathbf{z}^\star \in \mathbb{R}^N$ be a solution to the following optimisation problem:

$$\text{minimise:} \quad J(\mathbf{z}) := \frac{\|\mathbf{z}_{\Lambda_0}\|_2^2 + \|\mathbf{z}_{\Lambda_1}\|_2^2}{\|\mathbf{z}_{\Lambda_2}\|_2^2} \tag{27}$$

$$\text{subject to:} \quad \frac{\|\mathbf{z}_{\Lambda_1}\|_p^p + \|\mathbf{z}_{\Lambda_2}\|_p^p}{\|\mathbf{z}_{\Lambda_0}\|_p^p} \leq 1 \tag{28}$$

$$\|\mathbf{z}\|_2^2 = 1 \tag{29}$$

$$\text{and} \quad z_i \geq z_{i+1} \geq 0 \tag{30}$$

Then $\mathbf{z}^\star$ is piecewise flat, and has the form:

$$\mathbf{z}^\star = [\underbrace{\alpha,\ldots,\alpha}_{m}, \underbrace{\beta,\ldots,\beta}_{L}, \gamma, 0, \ldots, 0]^T \tag{31}$$

for some constants $\alpha > \beta > \gamma \geq 0$ and some L such that $k+1 \leq m+L \leq N$. Furthermore (28) holds with equality for $\mathbf{z}^\star$.

Proof: We first note that, due to the continuity of $J(\mathbf{z})$ and the compactness of the constraint set, an optimum $\mathbf{z}^\star$ is guaranteed to exist. Then we prove by contradiction that $\mathbf{z}^\star$ must have the claimed form.

- $\mathbf{z}^\star_{\Lambda_0}$ **is flat**. We know that

$$\|\mathbf{z}^\star_{\Lambda_0}\|_2 \geq m^{1/2-1/p} \|\mathbf{z}^\star_{\Lambda_0}\|_p$$

  with equality only if $z_i^\star = m^{-1/p} \cdot \|\mathbf{z}^\star_{\Lambda_0}\|_p$ for all $i \in \Lambda_0$, i.e. if $\mathbf{z}^\star_{\Lambda_0}$ is "flat". We define $\mathbf{z}'$ by $\mathbf{z}'_{\Lambda_1\cup\Lambda_2} = \mathbf{z}^\star_{\Lambda_1\cup\Lambda_2}$ and $z_i' = m^{-1/p} \cdot \|\mathbf{z}^\star_{\Lambda_0}\|_p$ Note that $\mathbf{z}'' = \mathbf{z}'/\|\mathbf{z}'\|_2$ is feasible. If $\mathbf{z}^\star_{\Lambda_0}$ was not flat, by the strict inequality above we would obtain $J(\mathbf{z}'') = J(\mathbf{z}') < J(\mathbf{z}^\star)$ which would contradict the fact that $\mathbf{z}^\star$ is an optimum. Hence $\mathbf{z}^\star_{\Lambda_0}$ must be flat.

- $\mathbf{z}^\star_{\Lambda_1}$ **is flat with all entries equal to** $z^\star_{k+1} = \|\mathbf{z}^\star_{\Lambda_2}\|_\infty$. By contradiction, assume that $z^\star_i \neq z^\star_{k+1}$ for some $i \in \Lambda_1$. Then, we can construct a $\mathbf{z}'$ with $\mathbf{z}'_{\Lambda_0 \cup \Lambda_2} = \mathbf{z}^\star_{\Lambda_0 \cup \Lambda_2}$ and $z'_i = z^\star_{k+1}$ for all $i \in \Lambda_1$. Again re-scale: $\mathbf{z}'' = \mathbf{z}'/\|\mathbf{z}'\|_2$. Thus $\mathbf{z}''$ is feasible and $J(\mathbf{z}'') = J(\mathbf{z}') < J(\mathbf{z}^\star)$. Hence $\mathbf{z}^\star_{\Lambda_1}$ must be flat with value $z^\star_{k+1}$.
- **Shape of** $\mathbf{z}^\star_{\Lambda_2}$. Let j be the smallest index such that $z^\star_j < z^\star_k$. Similarly let l be the largest index such that $z^\star_l > 0$. Suppose that $j \neq l$, otherwise we already have the form in (31). We can now construct a $\mathbf{z}'$ with non-negative, non-increasing entries such that $z'_i = z^\star_i$ for all $i \neq \{j, l\}$ with $\|\mathbf{z}'\|_p^p = \|\mathbf{z}^\star\|_p^p$ as follows. If $(z^\star_j)^p + (z^\star_l)^p \leq (z^\star_k)^p$ we set:

$$z'_l = 0 \text{ and } z'_j = \left((z^\star_j)^p + (z^\star_l)^p\right)^{1/p}$$

Otherwise we set:

$$z'_k = z^\star_k \text{ and } z'_l = \left((z^\star_j)^p + (z^\star_l)^p - (z^\star_k)^p\right)^{1/p}$$

Lemma 2 implies that $\|\mathbf{z}'_{\Lambda_2}\|_2 > \|\mathbf{z}^\star_{\Lambda_2}\|_2$, hence $J(\mathbf{z}') < J(\mathbf{z}^\star)$. Again we can re-scale to make the vector feasible. We can therefore conclude that $\mathbf{z}^\star_{\Lambda_2}$ can only have one element not equal to $z^\star_k$ or 0. This concludes the proof that $\mathbf{z}^\star$ must have the form in (31) with $\alpha \geq \beta > \gamma \geq 0$ and $k+1 \leq m+L \leq N$. Moreover by (28) we have

$$m \cdot \alpha^p = \|\mathbf{z}^\star_{\Lambda_0}\|_p^p \geq \|\mathbf{z}^\star_{\Lambda_1}\|_p^p + \|\mathbf{z}^\star_{\Lambda_2}\|_p^p \geq L \cdot \beta^p \geq (k+1-m) \cdot \beta^p > m \cdot \beta^p$$

hence $\alpha > \beta$.
- **Constraint** (28) **hold with equality for** $\mathbf{z}^\star$. Suppose that the left-hand side of (28) is strictly less than one for $\mathbf{z}^\star$. Since $\alpha > \beta$, we could then find $a < 1$ such that $\mathbf{z}'$ defined with $\mathbf{z}'_{\Lambda_1 \cup \Lambda_2} = \mathbf{z}^\star_{\Lambda_1 \cup \Lambda_2}$ and $\mathbf{z}'_{\Lambda_0} = a\mathbf{z}^\star_{\Lambda_0}$, properly re-scaled, simultaneously reduces the objective function (27) while still satisfying (28) and (30). Therefore (28) must hold with equality for any optimal $\mathbf{z}^\star$.

■

Lemma 3 implies that we only have to consider a relatively simple form for $\mathbf{z}$, which is parameterised by $\alpha > \beta > \gamma \geq 0$ and m, L, where $k - m + 1 \leq L \leq N - m$. Note that from Lemma 3 any zero elements in $\mathbf{z}$ can be removed without altering the optimal σ_k^2 by simply reducing the dimension N of the dictionary. In order to calculate the largest σ_k^2 we need to evaluate optimal values for α, β, γ, m and L. In fact we will see that we can ignore γ, which comes from the fact that the optimal constructions will correspond to m and N very large. The following lemma is expressed for $k = 2m$ but straightforward modifications would make it possible to handle arbitrary $k \geq 2m$.

Lemma 4 (Calculating the largest σ_{2m}^2): Consider $k = 2m < N$, $0 < p \leq 1$ and let η_p be the unique positive solution to (15). Let $\mathbf{z} \in \mathbb{R}^N$ be of the form (31) with $\alpha > \beta > \gamma \geq 0$ and $m + 1 \leq L \leq N - m$, and assume that $\mathbf{z}$ satisfies (28) with equality and (29). Then

$$\|\mathbf{z}_{\Omega_{2m}}\|_2^2 \geq \frac{2}{2-p}\eta_p. \tag{32}$$

If η_p is rational, equality is achieved for some $\mathbf{z}^\star_p$. Otherwise, the inequality can be replaced with $>$, but one can get arbitrarily close to the lower bound with appropriate choices of $k = 2m < N$ and $\mathbf{z}$ satisfying all the above conditions.

Proof: Define

$$L' := L + (\gamma/\beta)^p \tag{33}$$

$$\eta := m/L'. \tag{34}$$

Since $\gamma < \beta$, we have $L \leq L' < L+1$, and since $m+1 \leq L$ we have $0 < \eta < 1$. The ℓ^p failure equality constraint (28) reads $m\alpha^p = L\beta^p + \gamma^p = L'\beta^p$ hence

$$\beta = \eta^{1/p} \cdot \alpha < \alpha \tag{35}$$

Similarly by (29) we have $m\alpha^2 + L\beta^2 + \gamma^2 = 1$, and we let the reader check that this implies

$$m\alpha^2 + L'\beta^2 = 1 + (\gamma/\beta)^p \beta^2 - \gamma^2 \geq 1 \tag{36}$$

with equality when L' is integer (i.e. when $\gamma = 0$). Substituting $\beta = \eta^{1/p} \cdot \alpha$ and $L' = m/\eta$ in (36) we obtain:

$$m\alpha^2 \geq \left(1 + \eta^{2/p-1}\right)^{-1} \tag{37}$$

and it follows that

$$\|\mathbf{z}_{\Omega_{2m}}\|_2^2 = m\alpha^2 + m\beta^2 = m\alpha^2 \left(1 + \eta^{2/p}\right) \geq \frac{\left(1 + \eta^{2/p}\right)}{\left(1 + \eta^{2/p-1}\right)} \tag{38}$$

Differentiating the right-hand side and equating to zero, we observe that the zero of the derivative indeed yields a minimum, and we obtain that the value η_p that minimises the bound on $\|\mathbf{z}_{\Omega_{2m}}\|_2^2$ for $0 < \eta < 1$ satisfies (15). Substituting this back into (38) gives:

$$\|\mathbf{z}_{\Omega_{2m}}\|_2^2 \geq \frac{2}{2-p}\eta_p \tag{39}$$

Now that we have established the bound we discuss its tightness. First, one can check that for $0 < p \leq 1$, Equation (15) always has a unique solution in the region $0 < \eta_p < 1$, though the solution does not appear to have a general closed form. Then, notice that by continuity, the right-hand side in (38) can get arbitrarily close to the right-hand side in (39) by choosing η sufficiently close to η_p. Moreover, by the density of the rational numbers in $\mathbb{R}$, we can always find integers m and L such that m/L gets arbitrarily close to η_p. For such integers, setting $\gamma = 0$ (so that $L' = L$ and $\eta = m/L$), choosing α to reach equality in (37), and setting β according to (35) yields a vector $\mathbf{z}^\star$ for which $\|\mathbf{z}^\star_{\Omega_{2m}}\|_2^2$ is arbitrarily close to the lower bound. If η_p is rational then equality is actually achieved. If η_p is irrational, then equality cannot be achieved. ■

We are now able to state the proof of Theorem 3.

Proof: [Proof of Theorem 3] Consider a unit spectral norm dictionary $\mathbf{\Phi}$. Assume that $\mathbf{\Phi}$ is ℓ^p failing for some m-sparse vector. Then, by Proposition 1, there exists a minimally redundant row orthonormal (unit spectral norm) dictionary $\mathbf{\Phi}^\star \in \mathbb{R}^{(N-1)\times N}$ which is ℓ^p failing for some m-sparse vector such that

$$\sigma_{2m}^2(\mathbf{\Phi}) \leq \sigma_{2m}^2(\mathbf{\Phi}^\star)$$

By Proposition 2,

$$\sigma_{2m}^2(\mathbf{\Phi}^\star) = 1 - \|\mathbf{z}_{\Omega_{2m}}\|_2^2$$

where $\mathbf{z}$ is a unit norm vector which spans the null space $\mathcal{N}(\mathbf{\Phi}^\star)$. Since $\mathbf{\Phi}^\star$ is ℓ^p failing, $\mathbf{z}$ (after proper reindexing and taking the absolute value) satisfies the constraints (28), (29) and (30), therefore by Lemma 3 and Lemma 4,

$$\|\mathbf{z}_{\Omega_{2m}}\|_2^2 \geq \frac{2}{2-p}\eta_p. \tag{40}$$

We conclude that

$$\sigma_{2m}^2(\mathbf{\Phi}) \leq 1 - \frac{2}{2-p}\eta_p.$$

By contraposition, if $\sigma_{2m}^2(\mathbf{\Phi}) > 1 - \frac{2}{2-p}\eta_p$ then $\mathbf{\Phi}$ cannot be ℓ^p failing for any m-sparse vector. If η_p is irrational, the inequality in (40) can be replaced with $>$ hence it is sufficient to assume that $\sigma_{2m}^2(\mathbf{\Phi}) \geq 1 - \frac{2}{2-p}\eta_p$.

Conversely, by the above Propositions and Lemmatas, for every $\epsilon > 0$ there exists some $\mathbf{z}^\star$ satisfying the constraints (28), (29) and (30) for which

$$\|\mathbf{z}^\star_{\Omega_{2m}}\|_2^2 \leq \frac{2}{2-p}\eta_p + \epsilon, \tag{41}$$

yielding a (minimally redundant, row orthonormal) unit spectral norm dictionary $\mathbf{\Phi}_p^\star$ with

$$\sigma_{2m}^2(\mathbf{\Phi}_p^\star) \geq 1 - \frac{2}{2-p}\eta_p - \epsilon$$

which is ℓ^p failing for some m-sparse vector. If η_p is rational, this is true with equality for $\epsilon = 0$. ■

Let us proceed with the proof of Corollary 1.

Proof: [Proof of Corollary 1] Since $\mathbf{\Phi}$ is suitably re-scaled row orthonormal dictionary, $\mathbf{\Phi} = A \cdot \tilde{\mathbf{\Phi}}$ for some row orthonormal dictionary $\tilde{\mathbf{\Phi}}$ and some real constant $0 < A < \infty$. For $N - M < 2m \leq M$, since $\mathbf{\Phi}$ is a re-scaled version of $\tilde{\mathbf{\Phi}}$, by Remark 1 we have

$$\frac{1 - \sigma_{2m}^2(\tilde{\mathbf{\Phi}})}{1 + \sigma_{2m}^2(\tilde{\mathbf{\Phi}})} = \delta_{2m}(\tilde{\mathbf{\Psi}}_{2m}) \leq \delta_{2m}(\mathbf{\Phi}) < \frac{\eta_p}{2 - p - \eta_p}$$

with $\tilde{\mathbf{\Psi}}_{2m}$ the optimally re-scaled version of $\tilde{\mathbf{\Phi}}$ given by (12), hence

$$\sigma_{2m}^2(\tilde{\mathbf{\Phi}}) > 1 - \frac{2}{2-p}\eta_p$$

and we can apply Theorem 3 to conclude. ■

We conclude this section with the proof of Lemma 1.

Proof: [Proof of Lemma 1] Consider $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1) \in \mathbb{R}^N$ where $N = 2m+1$, $\mathbf{z}_0 \in \mathbb{R}^m$ is "flat" with entries $1/\sqrt{2m}$ and $\mathbf{z}_1 \in \mathbb{R}^{m+1}$ is "flat" with entries $1/\sqrt{2m+2}$. Check that $\mathbf{z}$ has non-increasing entries, is ℓ^2 normalised and satisfies the ℓ^p recovery condition for m-sparse vectors for $p = 0$ as well as for every $0 < p \leq 1$:

$$\|\mathbf{z}_0\|_p = m^{1/p-1/2} \cdot \|\mathbf{z}_0\|_2 = m^{1/p-1/2} \cdot \|\mathbf{z}_1\|_2 < (m+1)^{1/p-1/2} \cdot \|\mathbf{z}_1\|_2 = \|\mathbf{z}_1\|_p. \tag{42}$$

Let $\mathbf{\Phi}_1 \in \mathbb{R}^{(N-1)\times N}$ be a row orthornormal dictionary with null space spanned by $\mathbf{z}$: by the above properties, for every $0 \leq p \leq 1$, every m-sparse vector is recovered by the minimisation (4). By Proposition 2, for every k we have $\sigma_k^2(\mathbf{\Phi}_1) = 1 - \|\mathbf{z}_{\Omega_k}\|_2^2$, and in particular with $k := 2m$ and $k' := 2m+1$ we obtain

$$\sigma_{2m}^2(\mathbf{\Phi}_1) = \frac{1}{2m+2}; \qquad \sigma_{2m+1}^2(\mathbf{\Phi}_1) = 0.$$

Moreover, since $k \geq 2$ and $k' \geq 2$, by Remark 1 we have

$$\delta_{2m}(\mathbf{\Phi}_2) = \frac{2m+1}{2m+3}; \qquad \delta_{2m+1}(\mathbf{\Phi}_3) = 1;$$

with $\mathbf{\Phi}_2 = \mathbf{\Psi}_{2m}$ and $\mathbf{\Phi}_3 = \mathbf{\Psi}_{2m+1}$ the appropriately re-scaled dictionaries. ■

IV. Numerical studies of the σ^2_{2m} and δ_{2m} conditions

We now take a brief look at numerical solutions for values of σ^2_{2m} and δ_{2m} for which ℓ^p recovery can fail.

A. Large dimensional ℓ^p failing dictionaries

The analysis carried out so far, which relies on constructions for large dimensions m and N, shows that

$$\sup_{m,\mathbf{\Phi}} \sigma^2_{2m}(\mathbf{\Phi}) = 1 - \frac{2\eta_p}{2-p}$$
$$\inf_{m,\mathbf{\Phi}} \delta_{2m}(\mathbf{\Phi}) = \frac{\eta_p}{2-p-\eta_p}$$

where the supremum is over integers m and unit spectral norm ℓ^p failing dictionaries $\mathbf{\Phi}$, the infimum is over integers m and ℓ^p failing row orthonormal dictionaries $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ with $2m \leq M < N < M+2m$. This provides two curves $\sigma^2(p)$ and $\delta(p)$ for which there exists ℓ^p failing dictionaries with σ^2_{2m} above (respectively δ_{2m} below) or arbitrarily close to $\sigma^2(p)$ (resp. $\delta(p)$). To compute these curves we need to solve for η_p in Equation (15). For $p \in \{1, 2/3, 1/2\}$, this is a polynomial equation of degree $d = 2/p \in \{2,3,4\}$ which roots have algebraic expressions. In practice we rely on numerical solvers to compute η_p, $\sigma^2(p)$ and $\delta(p)$, which are displayed as a solid line on Figure 3 and Figure 4.

The work of [15], [16] showed that there is a whole family of sparsity measures including ℓ^p that span between ℓ^0 and ℓ^1, and that solving (4) for $p < 1$ could offer gradually superior performance to ℓ^1 recovery when p decreases. The results in [11] provided quantitative ℓ^p recovery conditions based on RIC. Here we see from Figure 4 that the offending RIC grows very gently as p shrinks. This implies that, at least in terms of worst case RIP analysis over all dictionaries, using a p slightly smaller than 1 does not provide a large benefit, and that one would need to rely on a $p \ll 1$ to expect a significant difference. However since solving (4) for $p < 1$ is non-convex such benefit will dependent on the specific choice of optimisation algorithm. For example we will see in the section V that iterative reweighted ℓ^1 techniques do not appear to provide uniform performance benefits beyond ℓ^1 minimisation.

These results may also seem at odds with a long history of empirical studies showing the benefits of ℓ^p optimisation for sparse recovery dating back to [13], however we note first that empirical results generally indicate an average performance bound rather than a uniform one and second the success of ℓ^p optimisation seems to be predominantly associated with sparsity problems with a range of coefficient sizes, such as Gaussian distributed sparse coefficients, where the ℓ^p algorithm is able to pick off the larger coefficients first. Note that successful recovery in ℓ^1 optimisation is only a function of the signs of the coefficients [12] and thus is unable to exploit differences in coefficient size.

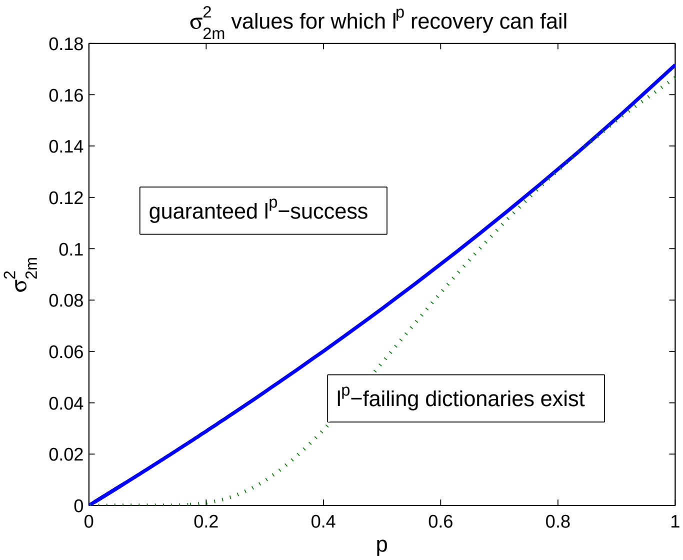

Fig. 3. A plot of the σ^2_{2m} values for which ℓ^p recovery with *unit spectral norm dictionaries* can fail (solid). This result is sharp in that for any (p, σ^2) strictly above the line, a dictionary with $\sigma^2_{2m} = \sigma^2$ is guaranteed to recover m-sparse representations by solving (4), while for any (p, σ^2) strictly below the line we can find a dictionary with $\sigma^2_{2m} = \sigma^2$ for which ℓ^p recovery will fail on at least one m-sparse vector. The dashed line corresponds to the values for the best failing 2×3 dictionaries calculated in section IV-B.

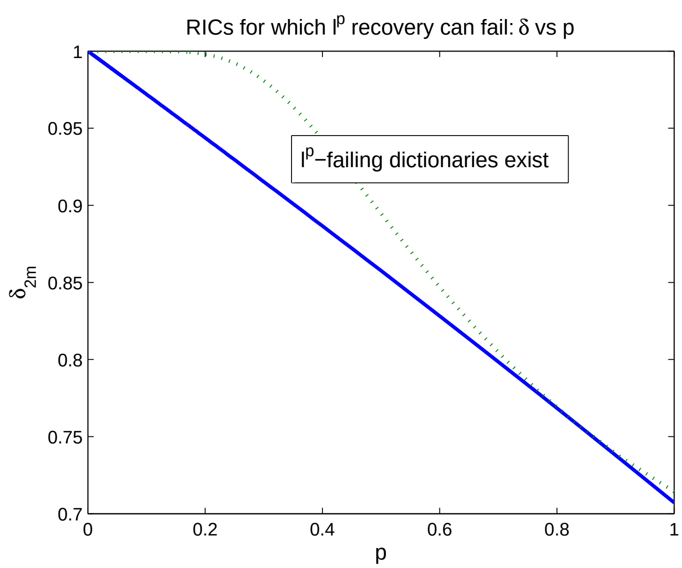

Fig. 4. A plot of the RIC values, δ for which ℓ^p recovery can fail (solid) for *any dictionary*. Strictly above the line, for any (p, δ), we can find a dictionary with $\delta_{2m} = \delta$ for which ℓ^p recovery will fail on at least one m-sparse vector. However, the result may not be sharp (except for the special case of row orthonormal dictionaries with $2m > N - M$, see Corollary 1) since δ_{2m} and σ^2_{2m} are only necessarily related through the inequality (13). Thus there may also exist failing dictionaries below the line. The dashed line corresponds to the RIC values for the re-scaled best failing 2x3 dictionaries in section IV-B.

B. Low dimensional examples

Although our arguments above require $N \to \infty$ in order to approach the bound, in fact, it is very easy to construct a specific low dimensional example that is very close to it. Consider a $\mathbf{\Phi} \in \mathbb{R}^{2\times 3}$ for which ℓ^p minimisation just fails in the 1-sparse case. Select:

$$\mathbf{z} = \frac{1}{\sqrt{1+2^{1-2/p}}} \cdot \begin{bmatrix} 1 \\ -2^{-1/p} \\ -2^{-1/p} \end{bmatrix} \tag{43}$$

and generate any $\mathbf{\Phi}$ such that $\mathbf{\Phi}^T$ is the orthogonal complement to $\mathbf{z}$. For example we can have:

$$\mathbf{\Phi} = \begin{bmatrix} \frac{1}{\sqrt{1+2^{2/p-1}}} & \frac{2^{1/p-1}}{\sqrt{1+2^{2/p-1}}} & \frac{2^{1/p-1}}{\sqrt{1+2^{2/p-1}}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \tag{44}$$

For the ℓ^1 case this gives:

$$\mathbf{z} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} \tag{45}$$

and

$$\mathbf{\Phi} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} & -\sqrt{3} \end{bmatrix} \tag{46}$$

The RIC, δ_2, that can fail for this (properly rescaled) low dimensional example is also plotted as a dashed line in Figure 4. It was simply computed by considering the three 2×2 submatrices of $\mathbf{\Phi}$ and computing their maximal and minimal singular values. Note that for ℓ^1 this is within 0.01 of the general condition for failure (due, no doubt, to the excellent engineering approximation of $\sqrt{2} \approx 3/2$: the offending $\mathbf{z}$ in (45) has the shape (31) for $m = 1, L = 2$, i.e. with $\eta = 1/2$, while the optimum is for $\eta_1 = \sqrt{2} - 1$). The value of p for which $\eta_p = 1/2$ is optimal can be found by numerically solving $(1/2)^{2/p} + 1 = 1/p$. This gives $p \approx 0.839$ for which the 2×3 construction is actually optimum. Note that the two curves in Figure 4 touch at this value of p.

V. Reweighted ℓ^1 implementations for ℓ^p optimisation

It is important to distinguish between optimisation functions and recovery algorithms. All the results in the previous sections have been derived for the recovery properties associated with the global minimum solutions for (4) without any regard for how these might be obtained. In practise, solving (4) for $p < 1$ is non-trivial. When $p < 1$ the cost function ceases to be convex and there are many local minima. One approach that has recently been proposed [9], [10], [4], [11] is to attempt to solve (4) by solving a sequence of reweighted ℓ^1 optimisation problems of the form:

$$\hat{\mathbf{y}}^{(n)} = \operatorname*{argmin}_{\mathbf{y}} \|\mathbf{W}_n \mathbf{y}\|_1 \text{ s.t. } \mathbf{\Phi y} = \mathbf{\Phi y}^\star, \tag{47}$$

where the initial weight matrix is set to the identity, $\mathbf{W}_1 = \mathbf{Id}$, and then subsequently $\mathbf{W}_n$ is selected as a diagonal positive definite weight matrix that is a (possibly iteration dependent) function of the previous solution vector,

$\mathbf{W}_n = f_n(\mathbf{y}^{(n-1)})$. At any step, the solutions to the convex optimisation problem (47) can be characterised by the necessary and sufficient property

$$\forall \mathbf{z} \in \mathcal{N}(\mathbf{\Phi})\backslash\{0\}, \qquad |\langle \mathbf{W}_n \mathbf{z}, \operatorname{sign}(\hat{\mathbf{y}}^{(n)})\rangle| \leq \|(\mathbf{W}_n \mathbf{z})_{\Gamma_n}\|_1 \tag{48}$$

where Γ_n denotes the set indexing the *zero* entries in $\hat{\mathbf{y}}^{(n)}$.

In [4], as an approximation to the ℓ^0 minimisation problem, the following reweighting function was proposed:

$$W_n(k,k) = \left(\epsilon_n + |y_k^{(n-1)}|\right)^{p-1} \tag{49}$$

for $p = 0$ and some small $\epsilon_n > 0$. Here $W_n(k,k)$ denotes the kth diagonal element of $\mathbf{W}_n$. In [11] the authors consider the same weighting function but including the full range of $0 \leq p < 1$. Note that the inclusion of the ϵ_n term is crucial as it keeps $W_n(k,k)$ bounded and ensures that a zero valued component y_i is able to become non-zero again at some subsequent iteration. Candès *et al.* [4] discuss using either a fixed ϵ_n or selecting it, while Foucart and Lai [11] argue that, at least in terms of the associated cost functions, letting $\epsilon_n \to 0$ converges to a solution for (4). It is also noted in [4] that there are various other reweighting strategies that could be deployed, some of which may not even be associated with a specific cost function.

A natural question to ask is: *what is the guaranteed performance of such algorithms?* In order to consider the widest possible set of reweighting schemes we define the following that we consider to encompass all 'reasonable' reweighting schemes.

Definition 2 (Admissible reweighting schemes): A reweighting scheme is considered to be admissible if, $\mathbf{W}_1 = \mathbf{Id}$ and if, for each n, there exists a $w^n_{\max} < \infty$ such that for all k, $0 \leq W_n(k,k) \leq w^n_{\max}$ and $W_n(k,k) = w^n_{\max} \Leftarrow \hat{y}_k^{(n-1)} = 0$.

The next two propositions shed some light on what performance guarantees we might expect from such schemes.

Proposition 3 (Iteratively reweighted ℓ^1 is not worse than ℓ^1): Let $\mathbf{\Phi}$ be an arbitrary dictionary and Ω an arbitrary support set. If ℓ^1 recovery is successful for all vectors with support set Ω, then recovery using any iteratively reweighted ℓ^1 algorithm with an admissible reweighting scheme is also successful for all vectors with support Ω.

Proof: Assume that Ω is a support set for which ℓ^1 is guaranteed to succeed: i.e., $\|\mathbf{z}_\Omega\|_1 < \|\mathbf{z}_{\Omega^c}\|_1, \forall \mathbf{z} \in \mathcal{N}(\mathbf{\Phi})\backslash\{0\}$. Since $\mathbf{W}_1 = \mathbf{Id}$, for any $\mathbf{y}^\star$ supported in Ω, $\hat{\mathbf{y}}^{(1)}$ is the ℓ^1 minimiser therefore $\hat{\mathbf{y}}^{(1)} = \mathbf{y}^\star$. As a result, $\Omega^c \subset \Gamma_1$, and for $k \in \Omega^c$, $W_2(k,k) = w^2_{\max}$, therefore

$$\forall \mathbf{z} \in \mathcal{N}(\mathbf{\Phi})\backslash\{0\}, \qquad |\langle \mathbf{W}_2 \mathbf{z}, \operatorname{sign}(\hat{\mathbf{y}}^{(1)})\rangle| \leq w^2_{\max} \cdot \|\mathbf{z}_\Omega\|_1 < w^2_{\max} \cdot \|\mathbf{z}_{\Omega^c}\|_1 \leq \|(\mathbf{W}_2 \mathbf{z})_{\Gamma_1}\|_1.$$

It follows that $\hat{\mathbf{y}}^{(2)} = \hat{\mathbf{y}}^{(1)}$ and iteratively one gets $\hat{\mathbf{y}}^{(n)} = \mathbf{y}^\star$ for all n. ■

Proposition 3 indicates that the reweighting strategy cannot damage an already successful solution. However we also have the following negative result.

Proposition 4 (Iteratively reweighted ℓ^1 is not uniformly better than ℓ^1): Let $\mathbf{\Phi} \in \mathbb{R}^{(N-1)\times N}$ be a minimally redundant dictionary of maximal rank $N-1$. Let Ω be a support set for which ℓ^1 recovery fails. Then *any* iteratively reweighted ℓ^1 algorithm with an admissible reweighting scheme will also fail for some vector $\mathbf{y}$ with support Ω.

Proof: Let $\mathbf{\Phi} \in \mathbb{R}^{(N-1)\times N}$ be a minimally redundant dictionary with maximal rank and let $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ be an arbitrary generator of its null space. Consider any set Ω for which ℓ^1 recovery can fail, i.e., $\|\mathbf{z}_\Omega\|_1 \geq \|\mathbf{z}_{\Omega^c}\|_1$. Let $\mathbf{y}^\star = \mathbf{z}_\Omega$. Because of the dimensionality of the null space, any representation satisfying $\mathbf{\Phi}\mathbf{y} = \mathbf{\Phi}\mathbf{y}^\star$ takes the form $\mathbf{y} = \mathbf{z}_\Omega - \alpha\mathbf{z} = (1-\alpha)\mathbf{z}_\Omega - \alpha\mathbf{z}_{\Omega^c}$. For any weight

$$\|\mathbf{W}_n\mathbf{y}\|_1 = |1-\alpha| \cdot \|\mathbf{W}_n\mathbf{z}_\Omega\|_1 + |\alpha| \cdot \|\mathbf{W}_n\mathbf{z}_{\Omega^c}\|_1, \alpha \in \mathbb{R}$$

hence there are only two possible unique solutions to (47), corresponding to $\alpha = 0$ and $\alpha = 1$. Since ℓ^1 fails to recover $\mathbf{y}^\star$, we have $\hat{\mathbf{y}}^{(1)} = -\mathbf{z}_{\Omega^c}$, therefore $\Omega \subset \Gamma_1$ and $\mathbf{W}_2(k,k) = w_{\max}^2$, $k \in \Omega$. [2] It follows that

$$|\langle \mathbf{W}_2\mathbf{z}, \text{sign}(\hat{\mathbf{y}}^{(1)})\rangle| \;\leq\; w_{\max}^2\|\mathbf{z}_{\Omega^c}\|_1 \;\leq\; w_{\max}^2\|\mathbf{z}_T\|_1 \;\leq\; \|(\mathbf{W}_2\mathbf{z})_{\Gamma_1}\|_1$$

and we obtain that $\hat{\mathbf{y}}^{(n)} = -\mathbf{z}_{\Omega^c}$ for all n. ■

Combining this with the results from section III immediately gives Theorem 2.

VI. Discussion

In this paper we have quantified values of the RIC, δ_{2m} for which there exist dictionaries where minimisation of (4) for some $0 < p \leq 1$ will fail to recover at least one m-sparse vector. This result is in some sense complementary to existing positive results [3], [11] and leaves limited room for improvement. Indeed for the special case of appropriately re-scaled row orthonormal dictionaries our negative result becomes sharp when $2m > N - M$.

On the other hand we have also shown that there exist minimally redundant row orthonormal dictionaries with RIC, δ_{2m} arbitrarily close to one for which ℓ^p recovery is successful for any p.[3] This should not be that surprising, RIP recovery conditions (be they for ℓ^1 or ℓ^p) come from a worst case analysis with respect to several parameters: worst case over all coefficients for a given sign pattern; worst case over all sign patterns for a given support; worst case over all supports of a given size; and worst case over all dictionaries with a given RIC. Our results emphasize the pessimism of such a worst case analysis.

In the context of compressed sensing [6], [3], there is also the desire to characterise the degree of undersampling (M/N) that is possible while still achieving exact recovery. Here RIP can be used to show that certain random matrices with high probability guarantee exact recovery with an undersampling of the order $(m/N)\log(N/m)$. However this result is indirect, firstly due to the worst case analysis discussed above and then secondly through the application of the concentration of measure [1]. A more direct approach, characterising the phase transition between exact recovery and undersampling for classes of random matrices, seems to provide a much clearer indication of the relationship between undersampling and recovery [8]. Of course, deriving expressions for such phase transitions

[2] If the solution to (47) is not unique then all values of α between 0 and 1 result in valid solutions and the algorithm has no means for determining the correct one. We therefore make the pessimistic assumption that the algorithm will select the incorrect representation associated with $\alpha = 1$.

[3] When the dictionary is not row orthonormal it is trivial to find such dictionaries by post-multiplying any ℓ^p successful dictionary with a matrix $A \in \mathbb{R}^{M\times M}$ that introduces the required ill-conditioning (i.e. $\mathbf{\Phi} \to A\mathbf{\Phi}$) to make $\delta_{2m} > 1 - \epsilon$. As the null space is unaffected by this action ℓ^p recovery is still maintained.

when $p \neq 1$ is likely to be a very challenging problem. Interestingly, the 'strong' phase transition of Donoho and Tanner [8] indicates that as $M/N \rightarrow 1$ most miminally redundant row orthonormal dictionaries will fail when $m/M \approx 0.18$. In contrast, our result for the ℓ^1-failing minimally redundant row orthonormal dictionary with the smallest RIC is associated with $m/M \rightarrow 1/(\sqrt{2}+2) \approx 0.29$ and so is clearly not indicative of the boundary behaviour.

Foucart and Lai [11] have also presented guaranteed recovery results for general ℓ^p minimisation with $0 < p \leq 1$. These results are couched in terms of δ_{2m+2} rather than δ_{2m} and are also explicitly dependent upon m. In contrast, the general result in Theorem 3 is independent of m though this could be refined to include m-dependence. Indeed for small m and p the positive result in [11] actually exceeds the negative bound computed from Theorem 3. However, we also note that for fixed p the m-dependent results rapidly converge to the m-independent result of $\delta_{2m+2} < 2(3-\sqrt{2})/7$, which is slightly weaker than their ℓ^1 recovery result since $\delta_{2m+2} \geq \delta_{2m}$. Theorem 3 seems to suggest that, at least in terms of worst case RIP analysis, there is limited benefit in reducing p a little below one.

Reducing $p < 1$ also introduces other issues. As the cost function is no longer convex the performance of the ℓ^p optimisation will be a function of the minimisation algorithm used. Our analysis of the iterative reweighted ℓ^1 algorithm (Theorem 2) shows that in terms of worst case RIP analysis there appears to be no gain in using this instead of the classical unweighted ℓ^1 minimisation.

Empirical evidence with iterative reweighting suggests that there can be substantial improvement over unweighted ℓ^1. However, while this might again be put down to the pessimistic nature of the worst case RIP analysis, we also suspect that the benefits of such algorithms stem from typically having a range of coefficient scales and that the performance of iterative reweighted ℓ^1 algorithms is probably highly coefficient dependent. Such non-uniform performance cannot be captured by a worst case performance analysis.

Although we have not explicitly considered it here, the RIP also plays a role in quantifying the robustness of ℓ^p recovery to observation noise [3], [11], i.e. when $\mathbf{x} = \mathbf{\Phi y} + \epsilon$. However, as noted here and in [11] exact recovery is independent of dictionary scaling, $\mathbf{\Phi} \rightarrow c\mathbf{\Phi}$, while robustness to noise is directly related to the scale of the dictionary. It is possible to define the error relative to the isometry constants as in [11], however it could be argued that a fairer measure of robustness would be in terms of absolute error when the dictionary is also constrained to have some physically reasonable property. For example, one might require that the dictionary or 'sensing matrix' cannot amplify observations, in other words $|||\mathbf{\Phi}||| \leq 1$. Interestingly, in this case, the notion of *asymmetric RIC* that we introduced in section II becomes the relevant measure. When viewed in this regard the existing robustness results for random matrices [3], [11] become significantly more pessimistic. This is because such matrices (e.g. random unit spectral norm orthoprojectors) typically shrink sparse vectors by a factor of $\sqrt{M/N}$ and to obtain an appropriate RIC requires re-scaling. However, this in turn implies that typically $|||\mathbf{\Phi}||| \approx \sqrt{N/M}$. Hence the robustness of *unit spectral norm* random matrices to observation error scales inversely proportional to the square root of the degree of undersampling.

We finally note that there are a couple of straightforward extensions that we have not pursued in order to keep the paper reasonably concise. First, it would be possible to extend the results in Remark 1 and Corollary 1 to include the factor A/B associated with non-tight frame bounds. Second, our main results are derived in terms of σ_k^2 and δ_k for $k = 2m$. However, there are a number of positive results based on RICs associated with larger index sets (as in [11]), $k > 2m$. Results similar to Lemma 4 and consequently Theorem 3 in terms of such sets should also be straightforward.

ACKNOWLEDGEMENTS

This research was partly supported by EPSRC grant D000246/1. The authors would like to thank Thomas Blumensath and Jared Tanner for many interesting discussions on ℓ^1 recovery and discussion of [3]. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

REFERENCES

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices." *To appear in Constructive Approximation*, 2008.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." *IEEE Trans. Info. Theory*, vol. 52, pp. 489–509, Feb 2006.

[3] E. Candès, "The Restricted Isometry Property and its implications for Compressed Sensing." *Comptes Rendus de l'Académie des Sciences, Paris, Série I*, 346, 589–592, 2008.

[4] E. J. Candés, M. B. Wakin and S. P. Boyd, "Enhancing sparsity by reweighted ell-1 minimization", to appear in J. Fourier Anal. Appl., 2008.

[5] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci Comp.*, vol. 20, no. 1, pp. 33-61, 1999.

[6] D. Donoho,"Compressed Sensing." *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp 1289-1306, April, 2006.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation from overcomplete dictionaries via ℓ^1 norm minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197-2002, Mar. 2002.

[8] D. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension", *to appear in Journal of the AMS*, 2008.

[9] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution", *In Proc. IEEE International Conference on Image Processing — ICIP'2005*, 2005.

[10] M. Figueiredo, J. Bioucas-Dias and R. Nowak, "Majorization-Minimization Algorithms for Wavelet-Based Image Restoration", *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[11] S. Foucart and M.-J. Lai,"Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$." *Submitted to Applied and Computational Harmonic Analysis*, 2008.

[12] J.-J. Fuchs, "On Sparse Representations in Arbitrary Redundant Bases", IEEE Trans. Inform. Theory, Vol. 50, No. 6, pp. 1341–1344, June 2004.

[13] I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm." *IEEE Trans. on Signal Processing*, vol. 45, pp. 600–616, March 1997.

[14] R. Gribonval and M. Nielsen, "Sparse decompositions in unions of bases." *IEEE Trans. Info. Theory*, vol. 49, no. 12, pp 3320-3325, Dec 2003.

[15] R. Gribonval and M. Nielsen, "On the strong uniqueness of highly sparse expansions from redundant dictionaries." *In Proc. Int Conf. Independent Component Analysis (ICA'04)*, Sep 2004.

[16] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure." *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp 335–355, May 2007. [Technical report October 2003].

[17] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals." *IEEE Trans. Info. Theory*, vol. 51, no. 3, pp. 1030-1051, Mar 2006.