On Lp minimisation, instance optimality, and restricted isometry constants for sparse approximation

Michael Davies, Rémi Gribonval

On Lp minimisation, instance optimality, and restricted isometry constants for sparse approximation

Michael Davies[(1)], Rémi Gribonval[(2)]

(1) IDCOM & Joint Research Institute for Signal and Image Processing
Edinburgh University, King's Buildings, Mayfield Road, Edinburgh, UK
(2) INRIA, Centre Inria Rennes - Bretagne Atlantique,
Campus de Beaulieu, F-35042 Rennes Cedex, France
mike.davies@ed.ac.uk, remi.gribonval@inria.fr

Abstract:

We extend recent results regarding the restricted isometry constants (RIC) and sparse recovery using ℓ^p minimisation. Here we consider the case of the sparse approximation of compressible rather than exactly sparse signals.

We begin by showing that the robust null space property used in [3] characterises the robustness of the estimation of compressible signals by ℓ^p mimisation for all ℓ^p norms, $0 < p \leq 1$ and not just ℓ^1, in the sense of instance optimality. Furthermore we show that this characterisation is in fact sharp in a particular sense. We are then able to simply extend the results in [6] to show that matrices that fail to exhibit good constants for instance optimality can be found with relatively small RICs.

1. Introduction

This paper considers conditions under which the solution $\hat{\mathbf{y}}$ of minimal ℓ^p quasi-norm, $0 < p \leq 1$, of an underdetermined linear system $\mathbf{x} = \mathbf{\Phi}\mathbf{y}$ is comparable with the best m-term approximation, $\mathbf{y}_m$. This problem is 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 data $\mathbf{y}$ that is assumed to be compressible (i.e. well approximated by some sparse vector $\mathbf{y}_m$). Most current guarantees on the performance of the extimate $\hat{\mathbf{y}}$ assume that the sensing matrix $\mathbf{\Phi}$ possesses a certain Restricted Isometry Constant (RIC), δ_{2m}. Recently in [6] the authors showed that the current best sufficient conditions [10], in terms of δ_{2m} for guaranteed recovery of exact m-sparse signals were close to optimal by finding a class of sensing matrix $\mathbf{\Phi}$ with small RIC for which sparse recovery can fail. Here we extend these results by considering the more general case of good sparse approximations instead of sparse representations, in the sense of instance optimality.

2. Notation

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 the equation $\mathbf{x} = \mathbf{\Phi}\mathbf{y}$. We will denote by $\|\mathbf{y}\|_p$ the ℓ^p sparsity measure defined as: $\|\mathbf{y}\|_p := \left(\sum_{j=1}^N |y_j|^p\right)^{1/p}$ where $0 < p \leq 1$. When $p = 0$, $\|\mathbf{y}\|_0$ denotes the ℓ^0 quasi-norm that counts the number of non-zero elements of $\mathbf{y}$. Everywhere in this paper, the notation $\|\cdot\|_p^p$ should be replaced with $\|\cdot\|_0$ when $p = 0$. 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.

3. Sparse recovery with ℓ^p minimisation

Throughout this paper we will consider signal estimates obtained as a solution of the following optimisation problem (which is non-convex for $0 \leq p < 1$):

$$\mathbf{y}_p^\star = \underset{\tilde{\mathbf{y}}}{\operatorname{argmin}} \|\tilde{\mathbf{y}}\|_p \text{ s.t. } \mathbf{\Phi}\tilde{\mathbf{y}} = \mathbf{\Phi}\mathbf{y}. \quad (1)$$

In the exact m-sparse case it has been shown [11] that if:

$$\|\mathbf{z}_\Omega\|_p < \|\mathbf{z}_{\Omega^c}\|_p \quad (2)$$

holds for all nonzero $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ then any vector $\mathbf{y}^\star$ whose support lies within Ω, is recovered as the *unique* solution of (1). If (2) holds for all $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ and all index sets Ω of size m, one says that $\mathbf{\Phi}$ satisfies the *null space property* (NSP) of order m, and a consequence is that any m-sparse vector $\mathbf{y}$ is recovered as the unique minimiser of (1). Furthermore the NSP is tight [12, 13].

When dealing with signals that are not exactly sparse one would like to control the approximation error of (1). To this end one defines the best m-term approximation error measured in the ℓ^p quasi-norm as:

$$\sigma_m(\mathbf{y})_p := \inf_{\|\tilde{\mathbf{y}}\|_0 \leq m} \|\mathbf{y} - \tilde{\mathbf{y}}\|_p \quad (3)$$

Following [5] one can ask whether it is possible to bound the approximation error of (1) in terms of the best m-term approximation error. If for all vectors $\mathbf{y}$ we have

$$\|\mathbf{y} - \mathbf{y}_p^\star\|_q \leq C \cdot \sigma_m(\mathbf{y})_q \quad (4)$$

then ℓ^p minimisation is "instance optimal of order m with constant C for the ℓ^q (quasi)norm" for the matrix $\mathbf{\Phi}$ [5].

It has been shown that instance optimality is related to a *robust null space property* (e.g. [3, 5]). The matrix $\mathbf{\Phi}$ satisfies the robust NSP of order m with constant $\rho < 1$ for the ℓ^p norm, or concisely $\mathbf{\Phi} \in \text{RobustNSP}(m, \rho, p)$, if

$$\|\mathbf{z}_\Omega\|_p^p < \rho\|\mathbf{z}_{\Omega^c}\|_p^p \tag{5}$$

for all nonzero $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ and all index sets Ω of size m. Note that by the results in [13] we have: *if $\mathbf{\Phi} \in \text{RobustNSP}(m, \rho, p)$ for some $0 < p \leq 1$, then $\mathbf{\Phi} \in \text{RobustNSP}(m, \rho, q)$ for all $0 \leq q \leq p$.*

If $\mathbf{\Phi} \in \text{RobustNSP}(m, \rho, p)$ then for all vectors $\mathbf{y}$ we have

$$\|\mathbf{y} - \hat{\mathbf{y}}\|_p^p \leq 2\frac{(1+\rho)}{(1-\rho)}\sigma_m(\mathbf{y})_p^p \tag{6}$$

whenever $\|\hat{\mathbf{y}}\|_p \leq \|\mathbf{y}\|_p$, and *in particular* for the minimum ℓ^p solution $\hat{\mathbf{y}} = \mathbf{y}_p^\star$ given by (1). The proof for $0 \leq p \leq 1$ directly follows the lines of the original proof for $p = 1$ in [3] and is also easily extended from $\|\cdot\|_p^p$ to general "f-norms" $\|\cdot\|_f$ as defined in [13]. It is omitted here. Condition (5) is also tight in the following sense.

Lemma 1 (Sharpness of the robust NSP). *Consider $0 < p \leq 1$ and $0 < \rho < 1$ such that*

$$3\rho + 1 \geq 2 \cdot \left(1 - \left(\frac{1-\rho}{2}\right)^{1/p}\right)^p. \tag{7}$$

If (5) fails for some $\mathbf{z} \in \mathcal{N}(\mathbf{\Phi})$ and some Ω of size (at most) m, then there exists $\mathbf{y}$ and $\hat{\mathbf{y}}$ with $\mathbf{\Phi}\mathbf{y} = \mathbf{\Phi}\hat{\mathbf{y}}$,

$$\|\hat{\mathbf{y}}\|_p \leq \|\mathbf{y}\|_p \text{ and } \|\mathbf{y} - \hat{\mathbf{y}}\|_p^p \geq 2\frac{(1+\rho)}{(1-\rho)}\sigma_m(\mathbf{y})_p^p. \tag{8}$$

The proof is given in the appendix. It is likely that the result can be extended to $p = 0$, and it is not known if the restriction (7) is necessary. Note that when the robust NSP with constant ρ is not satisfied, the Lemma does not rule out the fact that the (unique) ℓ^p minimiser $\mathbf{y}_p^\star$ might still always satisfy (6).

Similar conditions for instance optimality of order m were given in [5, Theorem 3.2] for general norms, and since their proof only uses the (quasi)triangle inequality they are easily extended to quasi-norms, such as ℓ^p quasi-norms for $0 \leq p \leq 1$. However, while the conditions in [5, Theorem 3.2] involve a robust NSP of order $2m$, with different constants for the necessary and the sufficient condition, Lemma 1 only involves the weaker robust NSP of order m, with matching constants.

Figure 1 illustrates the values (p, ρ) that satisfy (7). When $\rho < 1/3$, (7) fails for sufficiently small p, but when $\rho \geq 1/3$, it holds for any $0 < p \leq 1$.

4. Role of Restricted Isometry Constants

Using (1), particularly when $p = 1$, has become a popular mean of solving for sparse representations and approximations. An important characteristic of a matrix that has been used to guarantee good approximation performance in the

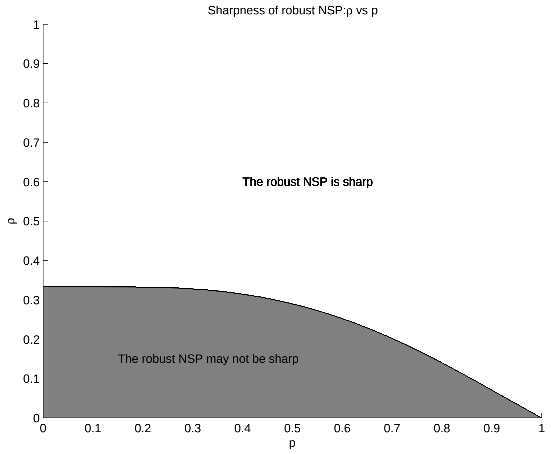

Figure 1: Overview of the values of (p, ρ) for which the robust NSP is known to be sharp according to Lemma 1.

form of (6) is the Restricted Isometry Property (RIP). For a matrix $\mathbf{\Phi}$ the restricted isometry constant (RIC), δ_k, is defined as the smallest number such that:

$$(1 - \delta_k) \leq \frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2} \leq (1 + \delta_k) \tag{9}$$

for every vector $\mathbf{y}$ and every index set Ω with $|\Omega| \leq k$.

If $\delta_{2m} < \sqrt{2} - 1$, then by [3, Lemma 2.2] the robust NSP (5) of order m holds with $\rho = \sqrt{2}\delta_{2m}(1 - \delta_{2m})^{-1} < 1$. The best known results for sparse approximation using (1) appear in [10].

Here we extend our negative results from [6]: we are interested in matrices $\mathbf{\Phi}$ with "small" RIC for which the robust NSP (5) for a given $0 < \rho < 1$ fails for some $z \in \mathcal{N}(\mathbf{\Phi})$. As in [6], we look for such failing matrices amongst the set of minimally redundant unit spectral norm matrices, *i.e.*, almost square matrices of size $(N-1) \times N$ with $\sup_{\mathbf{y} \neq 0} \|\mathbf{\Phi}\mathbf{y}\|_2^2/\|\mathbf{y}\|_2^2 = 1$. We exploit the fact [6, Remark 1] that the minimal squared singular value of any $2m$-submatrix of such $\mathbf{\Phi}$, defined as

$$\lambda_{2m}^2(\mathbf{\Phi}) := \min_{\substack{\mathbf{y}_\Omega \neq 0 \\ |\Omega| \leq 2m}} \frac{\|\mathbf{\Phi}\mathbf{y}_\Omega\|_2^2}{\|\mathbf{y}_\Omega\|_2^2}, \tag{10}$$

yields the RIC $\delta_{2m} = (1 + \lambda_{2m}^2)(1 - \lambda_{2m}^2)^{-1}$ for an appropriately rescaled matrix, and that λ_{2m}^2 can be completely characterised by the one-dimensional null space. Let $\mathbf{z} \in \mathbb{R}^N$ ($\|\mathbf{z}\|_2 = 1$) span the space $\mathcal{N}(\mathbf{\Phi})$, then: $\lambda_k^2(\mathbf{\Phi}) = 1 - \|\mathbf{z}_{\Omega_k}\|_2^2$ with Ω_k the index of the k largest components of $\mathbf{z}$. The problem of finding a matrix that fails the robust NSP (5) of order m while having maximal $\lambda_{2m}^2(\mathbf{\Phi})$ can therefore be transformed into a constrained optimisation problem as in [6]. Similarly the optimal null vector has a particularly simple form as indicated in the following lemmatas.

Lemma 2 (Shape of the optimal null vector $\mathbf{z}$). *Assume that $\rho > m/(N - m)$. Now consider $k \geq 2m$ and denote $\Lambda_0 := [\![1, m]\!]$, $\Lambda_1 := [\![m + 1, k]\!]$ and $\Lambda_2 := [\![k + 1, N]\!]$. Let $\mathbf{z}^\star \in \mathbb{R}^N$ be a solution to the following optimisation* *problem, with* $0 \le p \le 1$*:*

$$\text{minimise:} \quad J(\mathbf{z}) := \frac{\|\mathbf{z}_{\Lambda_0}\|_2^2 + \|\mathbf{z}_{\Lambda_1}\|_2^2}{\|\mathbf{z}_{\Lambda_2}\|_2^2} \tag{11}$$

$$\text{subject to:} \quad \frac{\|\mathbf{z}_{\Lambda_1}\|_p^p + \|\mathbf{z}_{\Lambda_2}\|_p^p}{\|\mathbf{z}_{\Lambda_0}\|_p^p} \le \frac{1}{\rho} \tag{12}$$

$$\|\mathbf{z}\|_2^2 = 1 \tag{13}$$

$$\text{and} \quad z_i \ge z_{i+1} \ge 0 \tag{14}$$

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

$$\mathbf{z}^\star = [\underbrace{\alpha, \ldots, \alpha}_{m}, \underbrace{\beta, \ldots, \beta}_{L}, \gamma, 0, \ldots, 0]^T \tag{15}$$

for some constants $\alpha \ge \beta > \gamma \ge 0$ *and some* L *such that* $k+1 \le m+L \le N$*. Furthermore (12) holds with equality for* $\mathbf{z}^\star$*.*

This lemma is identical to Lemma 3 in [6], except for the inclusion of ρ in (12) and the assumption $\rho > m/(N-m)$. The proof is essentially the same and is omitted here.
When $\rho \le m/(N-m)$ we have the following result.

Lemma 3 (Shape of the optimal null vector $\mathbf{z}$, trivial case)**.** *Suppose* $\rho \le m/(N-m)$ *then the solution to the optimisation problem* (11) - (14) *is given by* $z_i^\star = 1/\sqrt{N}$*.*

Proof. The vector with components $z_i^\star = 1/\sqrt{N}$ is the solution of the optimisation problem if we ignore the inequality constraint (12). However then we also have:

$$\frac{\|\mathbf{z}_{\Lambda_1}\|_p^p + \|\mathbf{z}_{\Lambda_2}\|_p^p}{\|\mathbf{z}_{\Lambda_0}\|_p^p} = \frac{N-m}{m} \le \frac{1}{\rho} \tag{16}$$

by our assumption on ρ, therefore the constraint (12) is automatically satisfied. □

From Lemma 3 it can be inferred with arguments similar to those developped in [6] that

$$\begin{aligned} \sup_{m,N,\mathbf{\Phi}} \lambda_{2m}^2(\mathbf{\Phi}) &= 1 - \inf_{m,N,\mathbf{z}} \|\mathbf{z}_{\Omega_{2m}}\|_2^2 \\ &= 1 - \inf_{m,N} \frac{2m}{N} \\ &= \frac{1-\rho}{1+\rho} \end{aligned} \tag{17}$$

where the supremum/infimum is over all integers m, N with $m/(N-m) \ge \rho$ and all minimally redundant unit spectral norm matrices, $\mathbf{\Phi}$ of size $(N-1)\times N$ / all vectors $\mathbf{z}$ satisfying (12)-(14). By [6, Remark 1], the smallest RIC for an appropriately rescaled sensing matrix $\mathbf{\Phi}$ under these constraints satisfies: $\inf_{m,N,\mathbf{\Phi}} \delta_{2m}(\mathbf{\Phi}) = \rho$.
If instead $\rho > m/(N-m)$ then the analysis directly follows that of [6] and the following result holds:

Lemma 4 (Calculating the largest λ_{2m}^2)**.** *Consider* $1 > \rho \ge m/(N-m)$*,* $k = 2m < N$*,* $0 < p \le 1$ *and let* η_p *be the unique positive solution to:*

$$\eta_p^{2/p} + \frac{2}{p} \cdot \rho^{2/p} \cdot \eta_p + (1 - \frac{2}{p})\rho^{2/p} = 0 \tag{18}$$

Let $\mathbf{z} \in \mathbb{R}^N$ *be of the form* (15) *with* $\alpha \ge \beta > \gamma \ge 0$ *and* $m+1 \le L \le N-m$*, and assume that* $\mathbf{z}$ *satisfies* (12) *with equality and* (13)*. Then*

$$\|\mathbf{z}_{\Omega_{2m}}\|_2^2 \ge \frac{2\eta_p}{2-p} \tag{19}$$

If η_p *is rational, equality is achieved for some* $\mathbf{z}_p^\star$*. 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.*

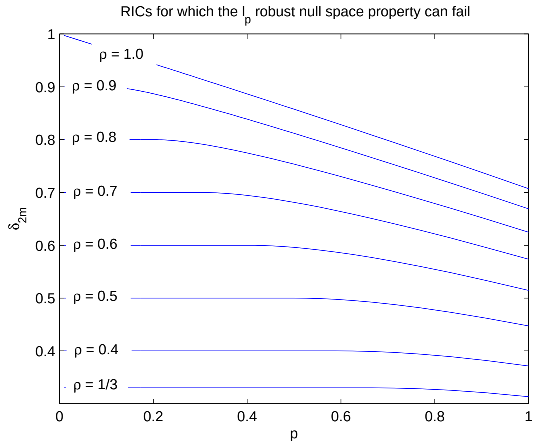

Figure 2: Curves demarking pairs of values $\{p, \delta_{2m}\}$ for which there exist matrices that fail the robust null space property of order m for different values of the constant ρ.

Although η_p does not have a nice analytic form for generic p, it can be computed numerically. Figure 2 displays (as a function of p, and for various constants ρ) the values of δ_{2m} for which it is possible to find sensing matrices that do not satisfy the robust NSP of order m . We need to be careful in our interpretation of these results, since by changing p we also change the quasi-norm in which the approximation performance is measured, but It is interesting to note that the effect of p is reduced as the value of ρ is reduced.

5. Discussion

Instance optimality expresses the robustness of ℓ^p minimisation when the signal to be estimated from noiseless incomplete measurements is not exactly sparse but only “compressible”. Robustness is quantified by the constant in (4), which should be as small as possible. In extending previous results on instance optimality of ℓ^p minimimisation from $p = 1$ to $0 \le p \le 1$, we highlighted the role of a robust null space property to characterise this constant.
Note that the robust NSP just depends on the null space of $\mathbf{\Phi}$, and is preserved when $\mathbf{\Phi}$ is replaced with $\mathbf{A\Phi}$ where $\mathbf{A}$ is any square invertible matrix. To the opposite, restricted isometry constants are “metric” properties of $\mathbf{\Phi}$ with respect to Euclidean norms, and it is only natural that the robust NSP can hold even for some matrices with large RIC. It should be straightforward to show that the robust NSP can hold at order m for any $\rho < 1$ even when δ_{2m} is arbitrarily close to one, in the spirit of [6, Lemma 1].
Phase transitions for the ℓ^1 recovery of exact-sparse signals from incomplete measurements have been characterised for various types of random sensing matrices as a function of the relative sparsity m/N and the compression ratio M/N for large N. The proofs involve a sub-

tle analysis of the geometry of random polytopes [9] associated with these matrices, but as we have seen the results are also sharply related with the phase transitions in terms of the satisfaction of a null space property of order m for $\rho = 1$. We believe that these results can be extended to phase transitions for the robust null space property for $p = 1$ and various values of $\rho < 1$. This should yield phase transitions associated with the robust estimation of compressible (rather than just exact-sparse) signals from incomplete measurements with ℓ^1 minimisation. It would be particularly interesting to see how the strong threshold for phase transitions [9] depends on ρ, and a related question is how the threshold associated to possible phase transitions for ℓ^p minimisation varies with p.

To conclude, let us note that while the results presented here rather highlight the limits of the RIC to predict the recovery of noiseless compressible signals, an important practical issue is robustness to noise, *i.e.*, when $\mathbf{x} = \mathbf{\Phi y} + \mathbf{e}$. In this case, robust recovery is guaranteed for small RIC [3], and this condition is likely to be necessary too. This will be the object of further investigations.

6. Acknowledgments

This work was supported in part by the European Union through the project SMALL (Sparse Models, Algorithms and Learning for Large-Scale data). The project SMALL acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 225913.

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.

A Proof of Lemma 1

Without loss of generality we assume that Ω indexes the m largest components of $\mathbf{z}$ and that $\|\mathbf{z}_\Omega\|_p^p = \rho \|\mathbf{z}_{\Omega^c}\|_p^p$ (otherwise we obtain a larger ρ' by changing either Ω or ρ, yielding a larger constant $2(1+\rho')/(1-\rho')$) in (8).

We define $\mathbf{y}_\Omega := \alpha \mathbf{z}_\Omega$, $\mathbf{y}_{\Omega^c} = \beta \mathbf{z}_{\Omega^c}$ and $\hat{\mathbf{y}} := \mathbf{y} + \mathbf{z}$ with $|\alpha| \geq |\beta|$ and $|\beta|^p := (1-\rho)/2$. We have $\mathbf{\Phi y} = \mathbf{\Phi} \hat{\mathbf{y}}$, and the m largest components of $\mathbf{y}$ are supported in Ω thus $\sigma_m(\mathbf{y})_p^p = \|\mathbf{y}_{\Omega^c}\|_p^p = \frac{1-\rho}{2} \cdot \|\mathbf{z}_{\Omega^c}\|_p^p$. As a result

$$\begin{aligned} \|\mathbf{y} - \hat{\mathbf{y}}\|_p^p &= \|\mathbf{z}\|_p^p = \|\mathbf{z}_\Omega\|_p^p + \|\mathbf{z}_{\Omega^c}\|_p^p \\ &= (\rho+1)\|\mathbf{z}_{\Omega^c}\|_p^p = 2\frac{(1+\rho)}{(1-\rho)}\sigma_m(\mathbf{y})_p^p \end{aligned}$$

To conclude we need to choose α so that $\|\hat{\mathbf{y}}\|_p^p \leq \|\mathbf{y}\|_p^p$. Since $\hat{\mathbf{y}}_\Omega = (\alpha - 1)\mathbf{z}_\Omega$ and $\hat{\mathbf{y}}_{\Omega^c} = (\beta-1)\mathbf{z}_{\Omega^c}$ we have

$$\begin{aligned} \|\hat{\mathbf{y}}\|_p^p &= |\alpha - 1|^p \cdot \|\mathbf{z}_\Omega\|_p^p + |\beta - 1|^p \cdot \|\mathbf{z}_{\Omega^c}\|_p^p \\ &= (\rho \cdot |\alpha - 1|^p + |\beta - 1|^p) \cdot \|\mathbf{z}_{\Omega^c}\|_p^p; \\ \|\mathbf{y}\|_p^p &= |\alpha|^p \cdot \|\mathbf{z}_\Omega\|_p^p + |\beta|^p \cdot \|\mathbf{z}_{\Omega^c}\|_p^p \\ &= (\rho \cdot |\alpha|^p + |\beta|^p) \cdot \|\mathbf{z}_{\Omega^c}\|_p^p \end{aligned}$$

and the constraint is

$$\rho \cdot (|\alpha|^p - |\alpha - 1|^p) \quad \geq \quad |\beta - 1|^p - |\beta|^p. \qquad (20)$$

Since $|\beta| = (\frac{1-\rho}{2})^{1/p} < 1/2$ the right-hand-side is positive. Therefore, if (20) is satisfied, its left-hand-side must also be positive and we must have $\alpha \geq 1/2$, which guarantees that $|\alpha| \geq |\beta|$. The left-hand side is maximum for $\alpha = 1$, and the right-hand-side is minimum for $\beta = +(\frac{1-\rho}{2})^{1/p}$ and we let the reader check that plugging these values of α and β in (20) yields the condition (7).

References:

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices." *Constructive Approximation*, 28(3), pp. 253-263, December 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] 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.

[5] A. Cohen, W. Dahmen and R. DeVore, "Compressed sensing and the best k-term approximation" *Preprint*, 2006.

[6] M. E. Davies and R. Gribonval, "Restricted Isometry Constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," to appear in IEEE Trans. Inf. Theory.

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

[8] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 norm minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197-2002, Mar. 2003.

[9] D. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension", *Journal of the AMS*, 22(1), pp. 1-53, January 2009.

[10] S. Foucart and M.-J. Lai,"Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$." *to appear in Applied and Computational Harmonic Analysis*, 2009.

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

[12] 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.

[13] 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].