A Numerical Exploration of Compressed Sampling Recovery

Charles Dossal
LaBag
Université Bordeaux 1
charles.dossal@math.u-bordeaux1.fr

Gabriel Peyré
Ceremade
Université Paris-Dauphine
gabriel.peyre@ceremade.dauphine.fr

Jalal Fadili
GREYC
ENSICAEN
jalal.fadili@greyc.ensicaen.fr

***Abstract*—This paper explores numerically the efficiency of ℓ^1 minimization for the recovery of sparse signals from compressed sampling measurements in the noiseless case. Inspired by topological criteria for ℓ^1-identifiability, a greedy algorithm computes sparse vectors that are difficult to recover by ℓ_1-minimization. We evaluate numerically the theoretical analysis without resorting to Monte-Carlo sampling, which tends to avoid worst case scenarios. This allows one to challenge sparse recovery conditions based on polytope projection and on the restricted isometry property.**

I. Compressed Sampling Recovery

Compressed sampling paradigm consists in acquiring a small number of linear measurements $y = Ax$, where $x \in \mathbb{R}^N$ is the high resolution signal to recover, and $y \in \mathbb{R}^P$ is the vector of measurements with $P \ll N$.

The resolution of the ill-posed inverse problem $y = Ax$ is stabilized by considering a matrix $A = (a_i)_{i=0}^{N-1} \in \mathbb{R}^{P \times N}$ that is drawn from a proper random ensemble. This article considers the case where the entries of A are drawn independently from a Gaussian variable of variance $1/P$.

For noiseless measurements $y = Ax$, the recovery of a sparse vector x is achieved by solving the convex program

$$x^\star = \underset{\tilde{x} \in \mathbb{R}^N}{\operatorname{argmin}} \|\tilde{x}\|_1 \quad \text{subj. to} \quad A\tilde{x} = y, \qquad (1)$$

where

$$\|\tilde{x}\|_1 = \sum_i |\tilde{x}_i|.$$

The vector x is said to be identifiable if $x^\star = x$. One usually seeks simple constraints on x that ensure stability. Of particular interest are constraints on the sparsity $\|x\|_0 = \#I(x)$ where the support of x is

$$I(x) = \{i \setminus x_i \neq 0\}.$$

With high probability on the sampling matrix A, compressed sampling theory [1], [2] shows that any vector satisfying

$$\|x\|_0 \leqslant \rho(P/N)P \qquad (2)$$

is identifiable for $\rho(\eta) > 0$ that increases slowly as the under-determinacy factor η decays to 0.

II. Sparse Recovery Criteria

Generic recovery criteria. Precise criteria for identifiability are obtained by considering the locations and signs of non-zero entries of x indexed by the support $I = I(x)$ of x. Such criteria use the interactions between the columns of $A_I = (a_i)_{i \in I}$ and the other ones $(a_i)_{i \notin I}$. Fuchs [3] proved that a sufficient condition for x to be identifiable is

$$F(x) = \max_{i \notin I} |\langle a_i,\, d(x)\rangle| < 1$$
$$\text{where} \quad d(x) = A_I (A_I^* A_I)^{-1} \operatorname{sign}(x_I),$$

see also Tropp [4] for a related result.

Topological recovery criteria. The centro-symmetric polytope $A(B_1)$ is the image of the ℓ^1 ball

$$B_1 = \{\tilde{x} \setminus \|\tilde{x}\|_1 \leqslant 1\}$$

and is the convex hull of $\{\pm a_i\}_i$. The $\|x\|_0$-dimensional facet $f_x \subset A(B_1)$ selected by x is the convex hull of $\{\pm a_i\}_{i \in I}$. Donoho [5] showed that

$$x \text{ is identifiable} \iff f_x \in \partial A(B_1) \qquad (3)$$

where $\partial A(B_1)$ is the boundary of the polytope $A(B_1)$. Dossal [6] shows that this topological condition is equivalent to having x as the limit of x_n where $F(x_n) < 1$.

Using (3), Donoho [5] determines, in the noiseless case $y = Ax$, a precise value for $\rho(\eta)$ in (2). For instance $\rho(1/2) \approx 0.089$ and $\rho(1/4) \approx 0.065$.

Restricted isometry criteria. Original works of Donoho [1], Candès, Romberg and Tao [2] focus on the stability of the compressed sampling decoder. Towards this goal, these authors introduced the restricted isometry property (RIP), which imposes that there exist constants $0 < \delta_s^{\min} \leqslant \delta_s^{\max} < 1$ such that for any $x \in \mathbb{R}^N$ with $\|x\|_0 \leqslant s$,

$$(1 - \delta_s^{\min})\|x\|^2 \leqslant \|Ax\|^2 \leqslant (1 + \delta_s^{\max})\|x\|^2. \qquad (4)$$

In the original work of Candès et al., the RIP is symmetric, and they require equal RIP constants, $\delta_s^{\min} = \delta_s^{\max} = \delta_s$. These authors proved that a small enough value of δ_{2s} ensures identifiability of all s-sparse vectors. This is achieved with high probability on A if $s \leqslant CP/\log(N/P)$, which corresponds to condition (2) with $\rho(\eta) \leqslant C/\log(\eta^{-1})$.

It turns out that the largest and smallest eigenvalues of the Gram matrix $A_I^* A_I$ do not deviate from 1 at the same rate. Using asymmetric RIP constants, Foucart and Lai [7] show that

$$(4\sqrt{2}-3)\delta_{2s}^{\min} + \delta_{2s}^{\max} < 4(\sqrt{2}-1) \qquad (5)$$

ensures identifiability of all s-sparse vector. Blanchard et al. [8] determine ρ_0 such that with high probability on A

$$\|x\|_0 \leqslant \rho_0(P/N)P \qquad (6)$$

ensures that condition (5) is in force. One necessarily has $\rho^0(\eta) \leqslant \rho(\eta)$ since condition (6) guarantees identifiability, but it also ensures a strong robustness to noisy measurements. This causes the constant ρ_0 to be quite small, and for instance $\rho_0(1/2) = 0.003$ and $\rho_0(1/4) = 0.0027$.

III. Interior facets and Non-identifiable Vectors

An heuristic for identifiability based on $1/\|d(x)\|$. From (3) we deduce that a non identifiable vector x corresponds to a facet f_x belonging to the interior of the polytope $A(B_1)$. The following property allows one to compute the distance of f_x to the center of the polytope.

Proposition 1. *For any vector x such that rank$(A_I) = |I|$, the distance from the facet f_x to 0 is $\frac{1}{\|d(x)\|}$.*

Proof: The distance of f_x to the 0 is the minimum of the distance between any hyperplan $\mathcal{H}$ containing f_x and 0. The definition $d(x) = A_I^* d(x) = \mathrm{sign}(x_I)$ implies that $\langle d(x),\, (\mathrm{sign}\, x_i) a_i\rangle = 1$ for all $i \in I$. The hyperplane

$$\mathcal{H}_x = \{u \setminus \langle d(x),\, u\rangle = 1\}$$

is such that for all $i \in I$, $(\mathrm{sign}\, x_i) a_i \in \mathcal{H}_x$ and thus $f_x \subset \mathcal{H}_x$. The distance between $\mathcal{H}_x$ and 0 is $1/\|d(x)\|$.
Let $\mathcal{H}_1 = \{u \setminus \langle c,\, u\rangle\}$ be anorther hyperplan such that $(\mathrm{sign}\, x_i) a_i \in \mathcal{H}_1$, for all $i \in I$. The distance between $\mathcal{H}_1$ and 0 is $\frac{1}{\|c\|}$. For all $i \in I$, we have $\langle c,\, a_i\rangle = \langle d(x),\, a_i\rangle$ and thus $\langle c - d(x),\, a_i\rangle = 0$. Since $d(x) \in \mathrm{span}(a_i)_{i\in I}$, $\langle c - d(x),\, d(x)\rangle = 0$ and then $\|c\|^2 = \|c - d(x)\|^2 + \|d(x)\|^2 > \|d(x)\|^2$, which completes the proof. ∎

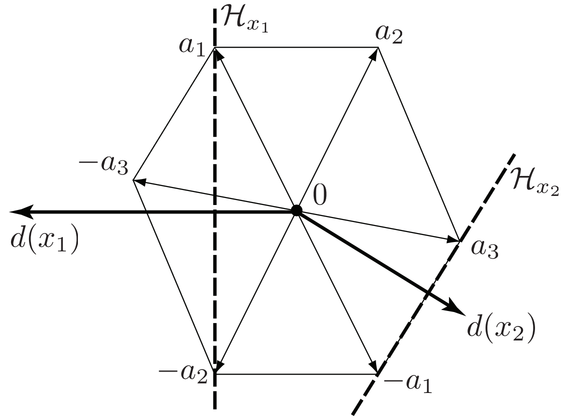

Figure 1. *Geometry of ℓ^1 recovery, for $N = 3$ and $P = 2$. The vector $x_1 = (2, -3, 0)$ is not identifiable because f_{x_1} is inside the polytope $A(B_1)$, and as a large $\|d(x_1)\|$. On the contrary, $x_2 = (-5, 0, 3)$ is identifiable because $f_{x_1} \in \partial A(B_1)$, and it has a small $\|d(x_1)\|$.*

Figure 1 illustrates this proposition for $P = 2$ dimensions. This property, together with condition (3), suggests that a vector x having a small value of $1/\|d(x)\|$ is more likely to be non identifiable.

Figure 2 estimates with a Monte-Carlo sampling the ratio of vectors that are identifiable, according to the sparsity $\|x\|_0$ and to a quantized value of $\|d(x)\|$. The curve parameterized by $\|d(x)\|$ exhibits a phase transition that is even sharper than the curve parameterized by sparsity (each dot on the curves accounts for 1000 andom realizations of the signal).

The numerical evidence provided by Figure 2 suggests that non-identifiable vectors might be found not just by increasing the sparsity of a given vector, but also by decreasing the value of $1/\|d(x)\|$.

An heuristic for sub-matrices conditioning based on $1/\|d(x)\|$. The following proposition shows that $1/\|d(x)\|$ is not only suitable to locate non-identifiable vectors, it is also a good proxy to extract sub-matrices of A that are ill-conditionned.

Proposition 2. *For any vector x such that rank$(A_I) = |I|$, one has*

$$\begin{cases} \delta_s^{\min} \geqslant 1 - s/\|d(x)\|^2, \\ \delta_s^{\max} \geqslant s/\|d(x)\|^2 - 1. \end{cases} \qquad (7)$$

Proof: One has $d(x) = (A_I^+)^* \,\mathrm{sign}(x_I)$, where A_I^+ is the pseudo-inverse of A_I, and hence

$$\lambda_{\min}((A_I^+)^*) \leqslant \frac{\|d(x)\|^2}{\|\,\mathrm{sign}(x_I)\|^2} \leqslant \lambda_{\max}((A_I^+)^*)$$

where $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ are the smallest and largest eigenvalues of B^*B. Since the eigenvalues of $A_I^* A_I$ are the inverse of those of $A_I^+(A_I^+)^*$ and

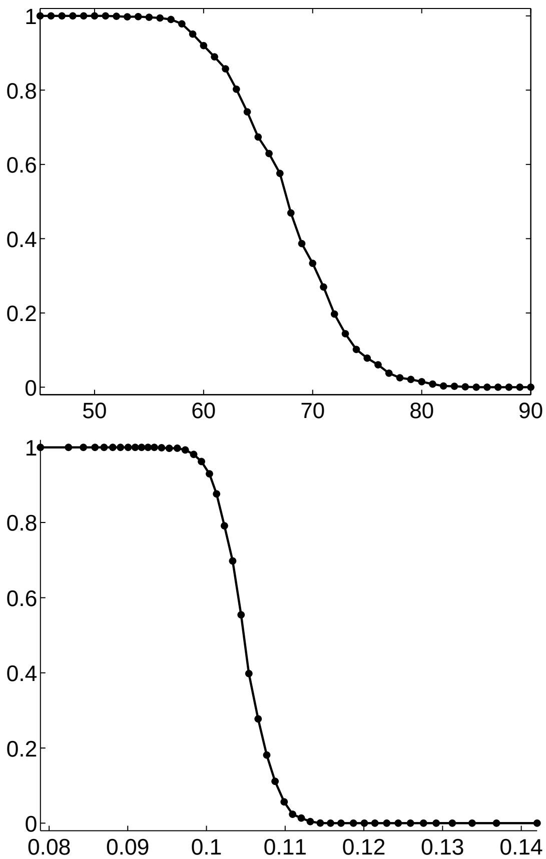

Figure 2. *Left: ratio of identifiable vectors as a function of $\|x\|_0$. Right: ratio of identifiable vectors as a function of $\|d(x)\|$.*

since $\|\operatorname{sign}(x_I)\|^2 = s$,

$$1 - \delta_s^{\min} \leqslant \lambda_{\min}(A_I) \leqslant \frac{s}{\|d(x)\|^2}$$
$$\frac{s}{\|d(x)\|^2} \leqslant \lambda_{\max}(A_I) \leqslant 1 + \delta_s^{\max},$$

which completes the proof. ■

Estimating a sharp lower bound on $\delta_s^{\max}$ (resp. $\delta_s^{\min}$) can thus be achieved by computing a vector x with $\|x\|_0 = s$ that maximizes (resp. minimizes) $1/\|d(x)\|$.

IV. Sparse Extensions

To build a vector x that is not identifiable, or a ill-conditionned sub-matrix A_I, we progressively increase the sparsity $\|x\|_0$ by extending it with additional non-zero entries. We consider a signed support extension $\tilde{x}$ of x, written as $\tilde{x} = x + \sigma\Delta_i$, where $\sigma \in \{+1, -1\}$, $i \notin I(x)$ and Δ_i is a Dirac vector. The extension of x to obtain $\tilde{x}$ increases by one the sparsity $\|x\|_0$, and we select carefully i and σ to maximize or minimize the variation of $1/\|d(x)\|$.

Approximate minimal and maximal worst extensions. Since $d(x) \in \operatorname{span}(a_j)_{j\in I(x)}$ and $I(x) \subset I(\tilde{x})$, we have

$$\langle d(x) - d(\tilde{x}),\, d(x)\rangle = 0,$$

hence

$$\|d(\tilde{x})\|^2 = \|d(x)\|^2 + \|d(x) - d(\tilde{x})\|^2.$$

Finding an extension that maximizes (resp. minimizes) $1/\|d(\tilde{x})\|$ is thus equivalent to minimizing (resp. maximizing) $\|d(x) - d(\tilde{x})\|$.

Introducing the dual vector

$$\tilde{a}_i \in \operatorname{span}((a_j)_{j\in I(x)} \cup a_i)$$

such that

$$\forall\, j \in I(x),\, \langle \tilde{a}_i,\, a_j\rangle = 0 \quad \text{and} \qquad \langle \tilde{a}_i,\, a_i\rangle = 1,$$

we have

$$d(\tilde{x}) - d(x) = -\tilde{a}_j(\langle d(x),\, a_j\rangle - \sigma),$$

which implies

$$\|d(x) - d(\tilde{x})\| = \|\tilde{a}_j\| |\langle d(x),\, a_j\rangle - \sigma| \;.$$

Computing $\|\tilde{a}_j\|$ for all possible $j \notin I(x)$ is computationally demanding since it requires to solve an over-determined system of linear equations for each j. We thus select an approximately optimal extension by maximizing or minimizing $|\langle d(x),\, a_j\rangle - \sigma|$ instead of $\|\tilde{a}_j\| |\langle d(x),\, a_j\rangle - \sigma|$.

The maximization (resp. minimization) of $1/\|d(x)\|$ is thus obtained through the extensions

$$\mathcal{E}^+(x) = x + \sigma^+\Delta_{i^+} \quad \text{and} \quad \mathcal{E}^-(x) = x + \sigma^-\Delta_{i^-}$$

$$\text{where} \begin{cases} i^+ = \underset{j\notin I(x)}{\operatorname{argmin}}\, |1 - |\langle d(x),\, a_j\rangle|| \\ i^- = \underset{j\notin I(x)}{\operatorname{argmax}}\, |\langle d(x),\, a_j\rangle| \end{cases} \tag{8}$$

$$\text{and} \begin{cases} \sigma^+ = \operatorname{sign}(\langle d(x),\, a_{i^-}\rangle), \\ \sigma^- = -\operatorname{sign}(\langle d(x),\, a_{i^+}\rangle). \end{cases} \tag{9}$$

Greedy minimal and maximal worst extensions. For each location $j \in \{0, \ldots, N-1\}$, starting from the initial 1-sparse vector $x_{j,0}^{\pm} = \Delta_j$, we compute iteratively two s-sparse worst case extensions as

$$\begin{cases} x_{j,s}^+ = \mathcal{E}^+(x_{j,s-1}^+) \\ x_{j,s}^- = \mathcal{E}^-(x_{j,s-1}^-). \end{cases} \tag{10}$$

V. Greedy Search for Non-identifiable Vectors

Proposition 1 suggests that the greedy extensions $x_{j,s}^-$ for varying j defined in (10) is likely to be difficult to identify.

Given $\eta = P/N \leqslant 1$, we use a dichotomy search on s to compute

$$s^\star(\eta, P) = \min\left\{s \setminus \exists j,\; x_{j,s}^- \text{ is not identifiable}\right\},$$

which is an empirical upper bound on the maximal allowable sparsity that guarantees identifiability.

The following table reports our numerical findings for $\eta = 1/4$, and compares this numerical evidence with the sharp theoretical bound of Donoho [5] $\rho(1/4) \sim 0.065$.

P	125	250	500	1000
$s^\star(1/4, P)$	10	20	42	79
$\lceil \rho(1/4)P \rceil$	9	17	33	65

For instance, with $N = 1000$ and $P = 250$, we are able to find a 20-sparse vector that is non-identifiable. In contrast, the Monte Carlo sampling displayed in Figure 2 does not reveal any non-identifiable vector for sparsity less than 54.

VI. Greedy Search for Ill-conditioned Sub-matrices

Empirical restricted isometry bounds. The extension $x_{j,s}^-$ defined in (10) is an s-sparse vector with a small value of $1/\|d(x)\|$. Proposition 2 suggests that its support $I = I(x_{j,s}^-)$ selects a Gram matrix $A_I^* A_I$ with a low smallest eigenvalue $\lambda_{\min}(A_I)$. Similarly, $I = I(x_{j,s}^+)$ can be used to find a sub-matrix $A_I^* A_I$ with a big largest eigenvalue $\lambda_{\max}(A_I)$.

We define empirical lower bounds on restricted isometry constants as

$$\left\{ \begin{array}{l} \tilde{\delta}_s^{\min} = \min\limits_{0 \leqslant j < N} 1 - \lambda_{\min}(A_{I(x_{j,s}^-)}) \\ \tilde{\delta}_s^{\max} = \max\limits_{0 \leqslant j < N} \lambda_{\max}(A_{I(x_{j,s}^+)}) - 1. \end{array} \right. \tag{11}$$

where $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ are the minimum and maximum eigenvalues of $B^* B$.

Figure 3 shows the numerical values of $\tilde{\delta}_s^{\min}$ and $\tilde{\delta}_s^{\max}$, and compare these bounds with more naive ones, obtained as follows.

– *Random sampling:* we use $K = 10^4$ sets $\{I_k\}_{k=0}^{K-1}$ of $\#I_k = s$ indexes, and use

$$\left\{ \begin{array}{l} \tilde{\delta}_{s,\text{rand}}^{\min} = \max_k 1 - \lambda_{\min}(A_{I_k}), \\ \tilde{\delta}_{s,\text{rand}}^{\max} = \max_k \lambda_{\max}(A_{I_k}) - 1, \end{array} \right.$$

as empirical lower bounds for $\delta_s^{\min}$ and $\delta_s^{\max}$.

– *Cone sampling:* for each $0 \leqslant j < N$, we select the s columns $(a_i)_{i \in I_j}$ of A that maximize $\langle a_i, a_j \rangle$. We use

$$\left\{ \begin{array}{l} \tilde{\delta}_{s,\text{cone}}^{\min} = \max_j 1 - \lambda_{\min}(A_{I_j}), \\ \tilde{\delta}_{s,\text{cone}}^{\max} = \max_j \lambda_{\max}(A_{I_j}) - 1, \end{array} \right.$$

as empirical lower bounds for $\delta_s^{\min}$ and $\delta_s^{\max}$.

This shows that our greedy algorithm is able to find ill-conditioned sub-matrices consistently better than simpler schemes.

We denote as x_s^- the vector reaching the empirical bounds (11), $\tilde{\delta}_s^{\min} = 1 - \lambda_{\min}(A_{I(x_s^-)})$. Figure 4 shows that the values of $1 - s/\|d(x_s^-)\|^2$ are close to the empirical restricted isometry constants $\tilde{\delta}_s^{\min}$. The same is true for the estimation of $\tilde{\delta}_s^{\max}$ using $s/\|d(x_s^+)\|^2 - 1$. This proves numerically that our heuristic (7) is accurate in practice.

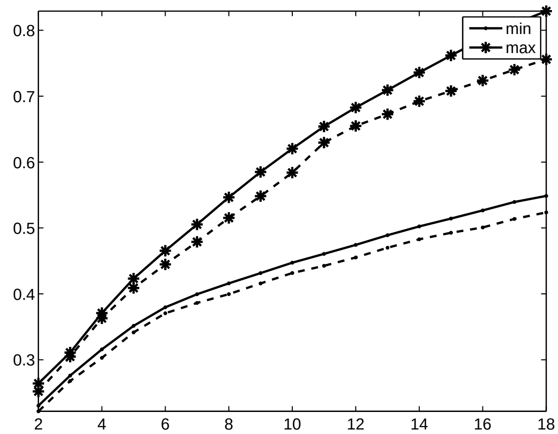

Figure 4. *Plain curves: values of $\tilde{\delta}_s^{\min}$ and $\tilde{\delta}_s^{\max}$ as a function of s. Dashed curves: values of $1 - s/\|d(x_s^-)\|^2$ and $s/\|d(x_s^+)\|^2 - 1$.*

Empirical sparsity bounds for restricted isometry condition. Given $\eta = P/N \leqslant 1$, we compute $s_0^\star(\eta, P)$, the minimum s for which our empirical estimates invalidate condition (5),

$$(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max} \geqslant 4(\sqrt{2} - 1)$$

Figure 5 shows our numerical estimation of the bound (5) for a varying s. The following table reports our numerical findings for $\eta = 1/4$, and compares this numerical evidence with the theoretical bound of Blanchard et al. [8] $\rho_0(1/4) \sim 0.0027$.

P	250	500	1000	2000
$s_0^\star(1/4, P)$	1	2	3	?
$\lceil \rho_0(1/4)P \rceil$	2	3	5	?

Conclusion

We have proposed in this paper new greedy algorithm to find sparse vector that are not identifiable and sub-matrices with a small number of columns that are ill-conditionned. This allows us to check numerically sparsity-based criteria for compressed sampling recovery based either on polytope projection or on restricted isometry constants.

References

[1] D. Donoho, “Compressed sensing,” *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[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, no. 2, pp. 489–509, 2006.

[3] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Trans. Info. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

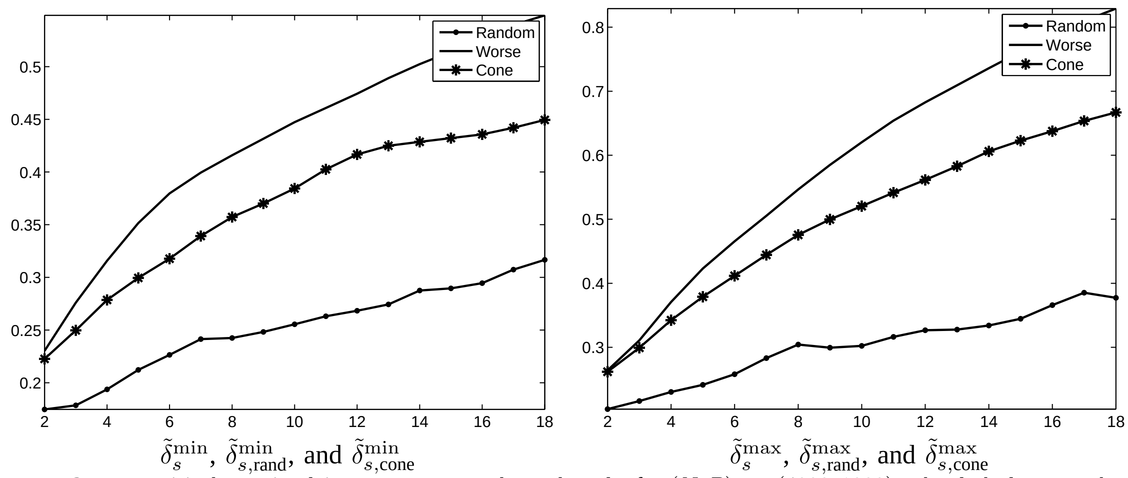

Figure 3. *Empirical Restricted isometry constants lower bounds, for* $(N, P) = (4000, 1000)$. *The dashed curves shows* $1 - s/\|d(x_s^-)\|^2$ *on the left and* $s/\|d(x_s^+)\|^2 - 1$ *on the right.*

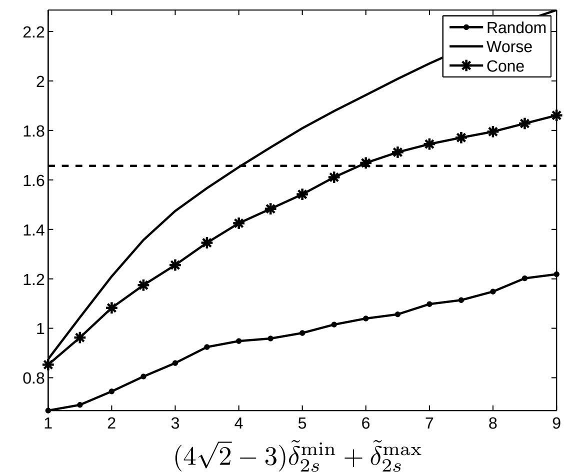

Figure 5. *Lower bound on condition* (5)*, for* $(N, P) = (4000, 1000)$*, the dashed line corresponds to* $y = 4(\sqrt{2} - 1)$.

[4] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Info. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[5] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[6] C. Dossal, "A necessary and sufficient condition for exact recovery by ℓ_1 minimization," *Preprint*, 2007.

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

[8] J. Blanchard, C. Cartis, and J. Tanner, "The restricted isometry property and ℓ^q-regularization: Phase transitions for sparse approximation," *Preprint*, 2009.