On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization

Anatoli Juditsky* Arkadi Nemirovski†

September 15, 2008

Abstract

We propose novel necessary and sufficient conditions for a sensing matrix to be "s-good" – to allow for exact ℓ_1-recovery of sparse signals with s nonzero entries when no measurement noise is present. Then we express the error bounds for imperfect ℓ_1-recovery (nonzero measurement noise, nearly s-sparse signal, near-optimal solution of the optimization problem yielding the ℓ_1-recovery) in terms of the characteristics underlying these conditions. Further, we demonstrate (and this is the principal result of the paper) that these characteristics, although difficult to evaluate, lead to verifiable sufficient conditions for exact sparse ℓ_1-recovery and to efficiently computable upper bounds on those s for which a given sensing matrix is s-good. We establish also instructive links between our approach and the basic concepts of the Compressed Sensing theory, like Restricted Isometry or Restricted Eigenvalue properties.

1 Introduction

In the existing literature on sparse signal recovery and Compressed Sensing (see [4-10,14-18] and references therein) the emphasis is on assessing sparse signal $x \in \mathbb{R}^n$ from an observation $y \in \mathbb{R}^k$ (in this context $k \ll n$):

$$y = Aw + \xi, \quad \|\xi\| \le \varepsilon, \tag{1.1}$$

where $\|\cdot\|$ is a given norm on $\mathbb{R}^k$, ξ is the observation error and $\varepsilon \ge 0$ is a given upper bound on the error magnitude, measured in the norm $\|\cdot\|$. One of the most popular (computationally tractable) estimators which is well suited for recovering sparse signals is the ℓ_1-*recovery* given by

$$\widehat{w} \in \operatorname{argmin}_z \{\|z\|_1 : \ \|Az - y\| \le \epsilon\}. \tag{1.2}$$

The existing Compressed Sensing theory focuses on this estimator and since our main motivation comes from the Compressed Sensing, we will also concentrate on this particular recovery. It is worth to mention that other closely related estimation techniques are used in statistical community, the most renown examples are "Dantzig Selector" (cf. [5]), provided by

$$\widehat{w}' \in \operatorname{argmin}_z \left\{\|z\|_1 : \ \|A^T(Az - y)\|_\infty \le \epsilon\right\}, \tag{1.3}$$

*LJK, Université J. Fourier, B.P. 53, 38041 Grenoble Cedex 9, France, Anatoli.Juditsky@imag.fr

†Georgia Institute of Technology, Atlanta, Georgia 30332, USA, nemirovs@isye.gatech.edu. Research of this author was supported by the Office of Naval Research grant #240667V.

and Lasso estimator, see [17, 4, 16], which under sparsity scenario exhibits similar behavior.

The theory offers strong results which state, in particular, that if w is s-sparse (i.e., has at most s nonzero entries) and A possesses a certain well-defined property, then the ℓ_1-recovery of x is close to x, provided the observation error ϵ is small. Some particularly impressive results make use of the Restricted Isometry property which is as follows: a $k \times n$ matrix A is said to possess the Restricted Isometry (RI(δ, m)) property with parameters $\delta \in (0, 1)$ and m, where m is a positive integer, if

$$\sqrt{1-\delta}\|x\|_2 \leq \|Ax\|_2 \leq \sqrt{1+\delta}\|x\|_2 \text{ for all } x \in \mathbb{R}^n \text{ with at most } m \text{ nonzero entries.} \tag{1.4}$$

For instance, the following result is well known ([10, Theorem 1.2] or [9, Theorem 4.1]): let $\|\cdot\|$ be the ℓ_2-norm, and let the sensing matrix A satisfy RI$(\delta, 2s)$-property with $\delta < \sqrt{2} - 1$. Then

$$\|\widehat{w} - w\|_1 \quad \leq \quad 2(1-\rho)^{-1}[\alpha\varepsilon\sqrt{s} + (1+\rho)\|w - w^s\|_1] \tag{1.5}$$

where $\alpha = \frac{2\sqrt{1+\delta}}{1-\delta}$, $\quad \rho = \frac{\sqrt{2}\delta}{1-\delta}$ and w^s is obtained from w by zeroing all but the s largest in absolute values entries. The conclusion is that when A is RI$(\delta, 2s)$ with $\delta < \sqrt{2}-1$, ℓ_1-recovery reproduces well signals with small s-tails (small $\|w - w^s\|_1$), provided that the observation error is small. Even more impressive is the fact that there are $k \times n$ sensing matrices A which possess, say, the RIP$(1/4, m)$-property for "large" m – as large as $O\left(k/\ln(n/k)\right)$ (this bound is tight). For instance, this is the case, with overwhelming probability, for matrices obtained by normalization (dividing columns by their $\|\cdot\|_2$-norms) of random matrices with i.i.d. standard Gaussian or ± 1 entries, as well as for normalizations of random submatrices of the Fourier transform or other orthogonal matrices.

On the negative side, random matrices are the only known matrices which possess the RI(δ, m)-property for so large m. For all known deterministic families of $k \times n$ matrices provably possessing the RI(δ, m)-property, one has $m = O(\sqrt{k})$ (see [12]), which is essentially worse than the bound $m = O(1)\left(k/\ln(n/k)\right)$ promised by the RI-based theory. Moreover, RI-property itself is "intractable" – the only currently available technique to verify the property for a $k \times n$ matrix amounts to test all its $k \times m$ submatrices. In other words, given a large sensing matrix A, one can never be sure that it possesses the RI(δ, m)-property with a given $m \gg 1$.

Certainly, the RI-property is not the only property of a sensing matrix A which allows to obtain good error bounds for ℓ_1-recovery of sparse signals. Two related characteristics are the Restricted Eigenvalue assumption, introduced in [4] and the Restricted Correlation assumption of [3], among others. However, they share with the RI-property not only the nice consequences as in (1.5), but also the drawback of being computationally intractable. To summarize our very restricted and sloppy description of the existing results on ℓ_1-recovery, neither Restricted Isometry, nor Restricted Correlation assumption and the like, do not allow to answer affirmatively the question whether for a *given sensing matrix* A, an accurate ℓ_1-recovery of sparse signals with a given number s of nonzero entries is possible.

Now, suppose we face the following problem: given a sensing matrix $\mathcal{A}$, which we are allowed to modify in certain ways to obtain a new matrix A, our objective is, depending on problem's specifications, either the maximal improvement, or the minimal deterioration of the sensing properties of A with respect to sparse ℓ_1-recovery. As a simple example, one can think, e.g., of a spatial or planar n-point grid E of possible locations of signal sources and an N-element grid R of possible locations of sensors A sensor in a given location measures a known, depending on the location, linear form of the signals emitted at the nodes of E, and the goal is to place a given number $k < N$ of sensors at the nodes of R in order to be able to recover, via the ℓ_1-recovery, all s-sparse signals. Formally speaking, we are given an $N \times n$ matrix $\mathcal{A}$, and our goal is to extract from it a $k \times n$ submatrix A which is *s-good* – such that whenever the true signal w in (1.1) is s-sparse and there is no observation error

($\xi = 0$), the ℓ_1-recovery (1.2) recovers w exactly. To the best of our knowledge, the only existing computationally tractable techniques which allow to approach such a synthesis problem are those based on *mutual incoherence*

$$\mu(A) = \max_{i \neq j} \frac{|A_i^T A_j|}{A_i^T A_i} \tag{1.6}$$

of a $k \times n$ sensing matrix A with columns A_i (assumed to be nonzero). Clearly, the mutual incoherence can be easily computed even for large matrices. Moreover, bounds of the same type as in (1.5) can be obtained for matrices with small mutual incoherence, as a matrix with the unit in $\|\cdot\|_2$-columns with mutual incoherence μ satisfies RI(δ, m) assumption (1.4) with $\delta = (m-1)\mu$. Unfortunately, the latter relation implies that μ should be very small to certify the possibility of accurate ℓ_1-recovery of non-trivial sparse signals, so that the estimates of a "goodness" of sensing for ℓ_1-recovery based on mutual incoherence are very conservative.

In this work, we aim at developing new *necessary and sufficient* characterization of s-goodness of a sensing matrix A. While these characterization is as difficult to verify as, say, the RI property, a good news is that this characterization yields *efficiently computable* upper and lower bounds on the values of s for which A is s-good. Aside of providing efficiently verifiable (although perhaps conservative) certificates for s-goodness, these bounds can be used when solving synthesis problems of the aforementioned type.

The overview of our main results is as follows.

1. We introduce (Section 2) a (difficult to compute) characteristic $\widehat{\gamma}_s(A)$ such that A is s-good if and only if $\widehat{\gamma}_s(A) < 1/2$. Thus, $\widehat{\gamma}_s(A)$ is fully responsible for ideal ℓ_1-recovery of s-sparse signals under *ideal* circumstances, when there is no observation error in (1.1) and (1.2) is solved to precise optimality.

2. In order to cope with the case of imperfect ℓ_1-recovery (nonzero observation error, nearly s-sparse true signal, near-optimal solving of (1.2)), we embed the characteristic $\widehat{\gamma}_s(A)$ into a single-parametric family of (still difficult to compute) characteristics $\widehat{\gamma}_s(A, \beta)$, $0 \leq \beta \leq \infty$, in such a way that $\widehat{\gamma}_s(A, \beta)$ is nonincreasing in β and is equal to $\widehat{\gamma}_s(A)$ for all large enough values of β. We then demonstrate (Section 3) that whenever $\beta < \infty$ is such that $\widehat{\gamma}_s(A, \beta) < 1/2$, the error of imperfect ℓ_1-recovery admits an explicit upper bound, similar in structure (although not identical to) the RI-based bound (1.5), expressed in terms of $\widehat{\gamma}_s(A, \beta)$, β, the measurement error ϵ, the "s-concentration" $\|w - w^s\|_1$ of the true signal w and the inaccuracy in solving (1.2).

3. In Section 4, we develop efficiently computable upper and lower bounds on $\widehat{\gamma}_s(A, \beta)$, thus arriving at efficiently verifiable *sufficient conditions* for s-goodness of A, and at efficiently computable upper bounds on the maximal s for which A is s-good. We demonstrate also that our verifiable sufficient conditions for s-goodness are less restrictive than those based on mutual incoherence. We present also "limits of performance" of our sufficient conditions for s-goodness: unfortunately, it turns out that these conditions, as applied to a $k \times n$ matrix A, cannot justify its s-goodness when $s > 2\sqrt{2k}$, unless A is "nearly square". While being much worse than the theoretically achievable, for appropriate A's, "$O(k/\ln(n/k))$-level of goodness", this "limit of performance" of our machinery nearly coincides with the best known "levels of goodness" of explicitly given individual $k \times n$ sensing matrices.

4. In Section 5, we investigate the implications of the RI property in our context. While these implications do not contribute to the "constructive" part of our results (since the RI property

is difficult to verify), they certainly contribute to better understanding of our approach and integrating it into the existing Compressed Sensing theory. The most instructive result of this Section is as follows: whenever A is, say, RI$(1/4, m)$ (so that the "true" level of s-goodness of A is $O(1)m$), our verifiable sufficient conditions do certify that A is $O(1)\sqrt{m}$-good – they guarantee "at least the square root of the true level of goodness".

5. Section 6 presents some very preliminary numerical illustrations of our machinery. These illustrations, in particular, present experimental evidence of how significantly this machinery can outperform the mutual-incoherence-based one – the only known to us existing computationally tractable way to certify goodness.

When this paper was finished, we became aware of two preprints [11, 1] which contain results closely related to some of those in our paper (we are greatly indebted to A. d'Aspremont for attracting our attention to these preprints). Specifically, in the relevant for us part of [11] the authors introduce a characteristic C_s of a $k \times n$ sensing matrix A with respect to a given sparsity level s: $C_s = \inf\{C' : \|x\|_1 \leq C'\|x - x^s\|_1 \, \forall x \in \mathrm{Ker}(A)\}$, where x^s is the best s-sparse $\|\cdot\|_1$-approximation of x, and demonstrate that this quantity is "fully responsible" for the possibility to recover x, given Ax, with $\|\cdot\|_1$-error bounded by constant times $\|x - x^s\|_1$; it should be stressed that the associated recovery is *not* the ℓ_1-one (and in fact is computationally intractable). Note that our $\widehat{\gamma}_s(A)$ is just a rescaling of C_s: $\widehat{\gamma}_s(A) = \frac{C_s - 1}{C_s}$, see Corollary 2.1. [11] contains also upper bounds on C_s for RI-matrices, similar to our bounds on $\widehat{\gamma}_s(A)$ in Proposition 5.1. The authors of [1] extract from [11] that when $C_s < 2$, the $\|\cdot\|_1$-error of the ℓ_1-recovery of x via Ax is bounded by $\frac{2C_s}{C_s - 2}\|x - x^s\|_1$ [1, Theorem 3], which is exactly the case of $\nu = 0$ of our Proposition 3.1. Besides this, [1] proposes an efficiently computable upper bound on C_s (and thus on $\widehat{\gamma}_s(A)$) based on semidefinite relaxation; this bound is completely different from ours, and it could be interesting to find out which one of these bounds is "stronger".

2 Characterizing s-goodness

2.1 Characteristics $\gamma_s(\cdot)$ and $\widehat{\gamma}_s(\cdot)$: definition and basic properties

The "minimal" requirement on a sensing matrix A to be suitable for recovering *s-sparse* signals (that is, those with at most s nonzero entries) via ℓ_1-minimization is as follows: whenever the observation y in (1.2) is noiseless and comes from an s-sparse signal w: $y = Aw$, w should be the unique optimal solution of the optimization problem in (1.2) where ϵ is set to 0. This observation motivates the following

Definition 2.1 *Let A be a $k \times n$ matrix and s be an integer, $0 \leq s \leq n$. We say that A is s-good, if for every s-sparse vector $w \in \mathbb{R}^n$, w is the unique optimal solution to the optimization problem*

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = Aw\}. \tag{2.7}$$

Let $s_*(A)$ be the largest s for which A is s-good; this is a well defined integer, since by trivial reasons every matrix is 0-good. It is immediately seen that $s_*(A) \leq \min[k, n]$ for every $k \times n$ matrix A.

From now on, $\|\cdot\|$ is the norm on $\mathbb{R}^k$ and $\|\cdot\|_*$ is its conjugate norm:

$$\|y\|_* = \max_v \left\{v^T y : \|v\| \leq 1\right\}.$$

We are about to introduce two quantities which are "responsible" for s-goodness.

Definition 2.2 *Let A be a $k \times n$ matrix, $\beta \in [0, \infty]$ and $s \leq n$ be a nonnegative integer. We define the quantities $\gamma_s(A, \beta)$, $\widehat{\gamma}_s(A, \beta)$ as follows:*
(i) *$\gamma_s(A, \beta)$ is the infinum of $\gamma \geq 0$ such that for every vector $z \in \mathbb{R}^n$ with s nonzero entries, equal to ± 1, there exists a vector $y \in \mathbb{R}^k$ such that*

$$\|y\|_* \leq \beta \ \& \ (A^T y)_i \begin{cases} = z_i, & z_i \neq 0 \\ \in [-\gamma, \gamma], & z_i = 0 \end{cases} ; \tag{2.8}$$

If for some z as above there does not exist y with $\|y\|_* \leq \beta$ such that $A^T y$ coincides with z on the support of z, we set $\gamma_s(A, \beta) = \infty$.
(ii) *$\widehat{\gamma}_s(A, \beta)$ is the infinum of $\gamma \geq 0$ such that for every vector $z \in \mathbb{R}^n$ with s nonzero entries, equal to ± 1, there exists a vector $y \in \mathbb{R}^k$ such that*

$$\|y\|_* \leq \beta \ \& \ \|A^T y - z\|_\infty \leq \gamma. \tag{2.9}$$

To save notation, we will skip indicating β when $\beta = \infty$, thus writing $\gamma_s(A)$ instead of $\gamma_s(A, \infty)$, and similarly for $\widehat{\gamma}_s$.

Several immediate observations are in order:
A. It is easily seen that the set of the values of γ participating in (i-ii) are closed, so that when $\gamma_s(A, \beta) < \infty$, then for every vector $z \in \mathbb{R}^n$ with s nonzero entries, equal to ± 1, there exists y such that

$$\|y\|_* \leq \beta \ \& \ (A^T y)_i \begin{cases} = z_i, & z_i \neq 0 \\ \in [-\gamma_s(A, \beta), \gamma_s(A, \beta)], & z_i = 0 \end{cases} ; \tag{2.10}$$

Similarly, for every z as above there exists $\widehat{y}$ such that

$$\|\widehat{y}\|_* \leq \beta \ \& \ \|A^T \widehat{y} - z\|_\infty \leq \widehat{\gamma}(A, \beta). \tag{2.11}$$

B. The quantities $\gamma_s(A, \beta)$ and $\widehat{\gamma}_s(A, \beta)$ are convex nonincreasing functions of β, $0 \leq \beta < \infty$. Moreover, from **A** it follows that for a given A, s and all large enough values of β one has $\gamma_s(A, \beta) = \gamma_s(A)$ and $\widehat{\gamma}_s(A, \beta) = \widehat{\gamma}_s(A)$.
C. Taking into account that the set $\{A^T y : \|y\|_* \leq \beta\}$ is convex, it follows that if $\gamma_s(A, \beta) < \infty$, then the vectors y satisfying (2.10) exist for every s-sparse vector z with $\|z\|_\infty \leq 1$, not only for vectors with exactly s nonzero entries equal to ± 1. Similarly, vectors $\widehat{y}$ satisfying (2.11) exist for all s-sparse z with $\|z\|_\infty \leq 1$. As a byproduct of these observations, we see that $\gamma_s(A, \beta)$ and $\widehat{\gamma}_s(A, \beta)$ are nondecreasing in s.

Our interest in the quantity $\gamma_s(\cdot, \cdot)$ stems from the following

Theorem 2.1 *Let A be a $k \times n$ matrix and $s \leq n$ be a nonnegative integer. Then A is s-good if and only if $\gamma_s(A) < 1$.*

Proof. a) Assume that A is s-good, and let us prove that $\gamma_s(A) < 1$. Let I be an s-element subset of the index set $\{1, ..., n\}$ and $\bar{I}$ be its complement, and let w be a vector supported on I and with nonzero w_i, $i \in I$. Then w should be the unique solution to the LP problem (2.7). ¿From the fact that w is an optimal solution to this problem it follows, by optimality conditions, that for certain y the function $f_y(x) = \|x\|_1 - y^T A^T x$ attains its minimum over $x \in \mathbb{R}^n$ at $x = w$, meaning that $0 \in \partial f_y(w)$, that is,

$$(A^T y)_i \begin{cases} = \mathrm{sign}(w_i), & i \in I \\ \in [-1, 1], & i \in \bar{I} \end{cases} ,$$

so that the LP problem

$$\min_{y,\gamma}\left\{\gamma : (A^Ty)_i \left\{\begin{array}{ll} = \mathrm{sign}(w_i), & i\in I \\ \in [-\gamma,\gamma], & i\in \bar{I}\end{array}\right.\right\} \tag{2.12}$$

has optimal value ≤ 1. Let us prove that in fact the optimal value is < 1. Indeed, assuming that the optimal value is exactly 1, there should exist Lagrange multipliers $\{\mu_i : i\in I\}$ and $\{\nu_i^\pm \ge 0 : i\in \bar{I}\}$ such that the function

$$\gamma + \sum_{i\notin I}\left[\nu_i^+[(A^Ty)_i-\gamma] + \nu_i^-[-(A^Ty)_i-\gamma]\right] - \sum_{i\in I}\mu_i\left[(A^Ty)_i - \mathrm{sign}(w_i)\right]$$

has unconstrained minimum in γ, y equal to 1, meaning that

$$\begin{array}{ll}(a) & \sum_{i\in\bar{I}}[\nu_i^+ + \nu_i^-] = 1,\\ (b) & \sum_{i\in I}\mu_i\mathrm{sign}(w_i) = 1,\\ (c) & Ad = 0, \text{ where } d\in\mathbb{R}^n \text{ with } d_i = \left\{\begin{array}{ll} -\mu_i, & i\in I\\ \nu_i^+ - \nu_i^-, & i\in\bar{I}.\end{array}\right.\end{array}$$

Now consider the vector $x_t = w + td$, where $t > 0$. This is a feasible solution to (2.7) due to (c); the $\|\cdot\|_1$-norm of this solution is

$$\sum_{i\in I}|w_i - t\mu_i| + t\sum_{i\in\bar{I}}|\nu_i^+ - \nu_i^-| \le \sum_{i\in I}|w_i - t\mu_i| + t$$

where the concluding inequality is given by (a) and the fact that $\nu_i^\pm \ge 0$. Since $w_i \ne 0$ for $i\in I$, for small positive t we have

$$\sum_{i\in I}|w_i - t\mu_i| = \sum_{i\in I}|w_i| - t\sum_{i\in I}\mu_i\mathrm{sign}(w_i) = \sum_{i\in I}|w_i| - t,$$

where the concluding equality is given by (b). We see that x_t is feasible for (2.7) and $\|x_t\|_1 \le \|w\|_1$ for all small positive t. Since w is the unique optimal solution to (2.7), we should have $x_t = w$, $t > 0$, which would imply that $\mu_i = 0$ for all i; but the latter is impossible by (b). Thus, the optimal value in (2.12) is < 1.

We see that whenever x is a vector with s nonzero entries, equal to ± 1, there exists y such that $(A^Ty)_i = x_i$ when $x_i \ne 0$ and $|(A^Ty)_i| < 1$ when $x_i = 0$ (indeed, in the role of this vector one can take the y-component of an optimal solution to the problem (2.12) coming from $w = x$), meaning that $\gamma_s(A) < 1$, as claimed.

b) Now assume that $\gamma_s(A) < 1$, and let us prove that A is s-good. Thus, let w be an s-sparse vector; we should prove that w is the unique optimal solution to (2.7). There is nothing to prove when $w = 0$. Now let $w \ne 0$, let s' be the number of nonzero entries of w, and I be the set of indices of these entries. By **C** we have $\gamma := \gamma_{s'}(A) \le \gamma_s(A)$, i.e., $\gamma < 1$. Recalling the definition of $\gamma_s(\cdot)$, there exists $y\in\mathbb{R}^k$ such that $(A^Ty)_i = \mathrm{sign}(w_i)$ when $w_i \ne 0$ and $|(A^Ty)_i| \le \gamma$ when $w_i = 0$. The function

$$f(x) = \|x\|_1 - y^T[Ax - Aw] = \sum_{i\in I}\left[|x_i| - \mathrm{sign}(w_i)(x_i - w_i)\right] + \sum_{i\notin I}\left[|x_i| - \gamma_i x_i\right],\ \gamma_i = (A^Ty)_i,\ i\notin I,$$

coincides with the objective of (2.7) on the feasible set of (2.7). Since $|\gamma_i| \le \gamma < 1$, this function attains its unconstrained minimum in x at $x = w$. Combining these two observations, we see that

$x = w$ is an optimal solution to (2.7). To see that this optimal solution is unique, let x' be another optimal solution to the problem. Then

$$0 = f(x') - f(w) = \sum_{i \in I} \underbrace{[|x'_i| - \mathrm{sign}(w_i)(x'_i - w_i) - |w_i|]}_{\geq 0} + \sum_{i \notin I} [|x'_i| - \gamma_i x'_i];$$

since $|\gamma_i| < 1$ for $i \notin I$, we conclude that $x'_i = 0$ for $i \notin I$. This conclusion combines with the relation $Ax' = Aw$ to imply the required relation $x' = w$, due to the following immediate observation:

Lemma 2.1 *If $\gamma_s(A) < 1$, then every $k \times s$ submatrix of A has trivial kernel.*

Proof. Let I be the set of column indices of a $k \times s$ submatrix of A. If this submatrix has a nontrivial kernel there exists a nonzero s-sparse vector $z \in \mathbb{R}^n$ such that $Az = 0$. Let I be the support set of z. By **A**, there exists a vector $y \in \mathbb{R}^k$ such that $(A^T y)_i = \mathrm{sign}(z_i)$ whenever $i \in I$, that is

$$0 = y^T Az = \sum_{i: z_i \neq 0} (A^T y)_i z_i = \|z\|_1,$$

which is impossible. ∎

Theorem 2.1 explains the importance of the characteristic $\gamma_s(\cdot)$ in the context of ℓ_1-recovery. However, it is technically more convenient to deal with the quantity $\widehat{\gamma}_s(\cdot)$. We shall address this issue in an instant, and for the time being we take note of the following result:

Proposition 2.1 *For every $\beta \in [0, \infty]$ one has*

$$\begin{array}{ll} (a) & \gamma := \gamma_s(A, \beta) < 1 \Rightarrow \widehat{\gamma}_s\left(A, \frac{1}{1+\gamma}\beta\right) = \frac{\gamma}{1+\gamma} < 1/2; \\ (b) & \widehat{\gamma} := \widehat{\gamma}_s(A, \beta) < 1/2 \Rightarrow \gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\beta\right) = \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1. \end{array} \tag{2.13}$$

Proof. Let $\gamma := \gamma_s(A, \beta) < 1$. By definition it means that for every vector $z \in \mathbb{R}^n$ with s nonzero entries, equal to ± 1, there exists y, $\|y\|_* \leq \beta$, such that $A^T y$ coincides with z on the support of z and is such that $\|A^T y - z\|_\infty \leq \gamma$. Given z, y as above and setting $y' = \frac{1}{1+\gamma} y$, we get $\|y'\|_* \leq \frac{1}{1+\gamma}\beta$ and

$$\|A^T y' - z\|_\infty \leq \max\left[1 - \frac{1}{1+\gamma}, \frac{\gamma}{1+\gamma}\right] = \frac{\gamma}{1+\gamma}.$$

Thus, for every vector z with s nonzero entries, equal to ± 1, there exists y' such that $\|y'\|_* \leq \frac{1}{1+\gamma}\beta$ and $\|A^T y' - z\|_\infty \leq \frac{\gamma}{1+\gamma}$, meaning that $\gamma := \gamma_s(A, \beta) < 1$ implies

$$\widehat{\gamma}_s\left(A, \frac{1}{1+\gamma}\beta\right) \leq \frac{\gamma}{1+\gamma} < 1/2. \tag{2.14}$$

Now assume that $\widehat{\gamma} := \widehat{\gamma}_s(A, \beta) < 1/2$. For an s-element subset I of the index set $\{1, ..., n\}$, let

$$\Pi_I = \left\{u \in \mathbb{R}^n : \text{ exists } y \in \mathbb{R}^k : \|y\|_* \leq \beta,\ (A^T y)_i = u_i \text{ for } i \in I,\ |(A^T y)_i| \leq \widehat{\gamma} \text{ for } i \in \bar{I}\right\},$$

where $\bar{I}$ is the complement of I. It is immediately seen that Π_I is a closed and convex set in $\mathbb{R}^n$. We claim that Π_I contains the centered at the origin $\|\cdot\|_\infty$-ball B of the radius $1 - \widehat{\gamma}$. Using this fact we conclude that for every vector z supported on I with entries z_i, $i \in I$, equal to ± 1, there exists an $u \in \Pi_I$ such that $u_i = (1 - \widehat{\gamma}) z_i$, $i \in I$. Recalling the definition of Π_I, we conclude that there exists

y with $\|y\|_* \le (1-\widehat{\gamma})^{-1}\beta$ such that $(A^Ty)_i = (1-\widehat{\gamma})^{-1}u_i = z_i$ for $i \in I$ and $|(A^Ty)_i| \le (1-\widehat{\gamma})^{-1}\widehat{\gamma}$ for $i \notin I$. Thus, the validity of our claim would imply that

$$\widehat{\gamma} := \widehat{\gamma}_s(A,\beta) < 1/2 \Rightarrow \gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\beta\right) \le \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1. \tag{2.15}$$

Let us prove our claim. Observe that by definition Π_I is the direct product of its projection Q on the plane $L_I = \{u \in \mathbb{R}^n : u_i = 0, i \notin I\}$ and the entire orthogonal complement $L_I^\perp = \{u \in \mathbb{R}^n : u_i = 0, i \in I\}$ of L_I; since Π_I is closed and convex, so is Q. Now, L_I can be naturally identified with $\mathbb{R}^s$, and our claim is exactly the statement that the image $\bar{Q} \subset \mathbb{R}^s$ of Q under this identification contains the centered at the origin $\|\cdot\|_\infty$ ball B_s, of the radius $1-\widehat{\gamma}$, in $\mathbb{R}^s$. Assume that it is not the case. Since $\bar{Q}$ is convex and $B_s \not\subset \bar{Q}$, there exists $v \in B_s \backslash \bar{Q}$, and therefore there exists a vector $e \in \mathbb{R}^s$, $\|e\|_1 = 1$ such that $e^Tv > \max_{v' \in \bar{Q}} e^Tv'$ (recall that Q, and thus $\bar{Q}$, is both convex and closed). Now let $z \in \mathbb{R}^n$ be the s-sparse vector supported on I such that the entries of z with indices $i \in I$ are the signs of the corresponding entries in e. By definition of $\widehat{\gamma} = \widehat{\gamma}_s(A,\beta)$, there exists $y \in \mathbb{R}^k$ such that $\|y\|_* \le \beta$ and $\|A^Ty - z\|_\infty \le \widehat{\gamma}$; recalling the definition of Π_I and $\bar{Q}$, this means that $\bar{Q}$ contains a vector $\bar{v}$ with $|\bar{v}_j - \mathrm{sign}(e_j)| \le \widehat{\gamma}$, $1 \le j \le s$, whence $e^T\bar{v} \ge \|e\|_1 - \widehat{\gamma}\|e\|_1 = 1-\widehat{\gamma}$. We now have

$$1-\widehat{\gamma} \ge \|v\|_\infty \ge e^Tv > e^T\bar{v} \ge 1-\widehat{\gamma},$$

where the first $\ge$ is due to $v \in B_s$, an $>$ is due to the origin of e. The resulting inequality is impossible, and thus our claim is true.

We have proved the relations (2.14), (2.15) which are slightly weakened versions of (2.13.*a*-*b*). It remains to prove is that the inequalities $\le$ in the conclusions of (2.14), (2.15) are in fact equalities. This is immediate: assume that under the premise of (2.13.*a*) we have

$$\widehat{\gamma} := \widehat{\gamma}_s\left(A, \frac{1}{1+\gamma}\beta\right) < \gamma_+ := \frac{\gamma}{1+\gamma}.$$

When applying (2.15) with β replaced with $\frac{1}{1+\gamma}\beta$, we get

$$\gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\left[\frac{1}{1+\gamma}\beta\right]\right) \le \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < \frac{\gamma_+}{1-\gamma_+} = \gamma. \tag{2.16}$$

At the same time, $\frac{1}{1-\widehat{\gamma}}\frac{1}{1+\gamma} < \frac{1}{1-\gamma_+}\frac{1}{1+\gamma} = 1$ due to $\widehat{\gamma} < \gamma_+$; since $\gamma_s(A,\cdot)$ is nonincreasing by **B**, we see that

$$\gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\left[\frac{1}{1+\gamma}\beta\right]\right) \ge \gamma_s(A,\beta),$$

and thus (2.16) implies that $\gamma_s(A,\beta) < \gamma$, which contradicts the definition of γ. Thus, the concluding $\le$ in (2.14) is in fact equality. By completely similar argument, so is the concluding $\le$ in (2.15). ∎

2.2 Equivalent representation of $\widehat{\gamma}_s(A)$

According to Proposition 2.1, the quantities $\gamma_s(\cdot)$ and $\widehat{\gamma}(\cdot)$ are tightly related. In particular, the equivalent characterization of s-goodness in terms of $\widehat{\gamma}_s(A)$ reads as follows:

$$A \text{ is } s\text{-good} \;\Leftrightarrow\; \widehat{\gamma}_s(A) < 1/2.$$

In the sequel, we shall heavily utilize an equivalent representation $\widehat{\gamma}_s(A,\beta)$ which, as we shall see in Section 4, has important algorithmic consequences. The representation is as follows:

Theorem 2.2 *Consider the polytope*

$$P_s = \{u \in \mathbb{R}^n : \ \|u\|_1 \le s, \ \|u\|_\infty \le 1\}.$$

One has

$$\widehat{\gamma}_s(A,\beta) = \max_{u,x} \left\{u^T x - \beta\|Ax\| : \ u \in P_s, \ \|x\|_1 \le 1\right\}. \tag{2.17}$$

In particular,

$$\widehat{\gamma}_s(A) = \max_{u,x} \left\{u^T x : \ u \in P_s, \ \|x\|_1 \le 1, \ Ax = 0\right\}. \tag{2.18}$$

Proof. By definition, $\widehat{\gamma}_s(A,\beta)$ is the smallest γ such that the closed convex set $C_{\gamma,\beta} := A^T B_\beta + \gamma B$, where $B_\beta = \{w \in \mathbb{R}^k : \ \|w\|_* \le \beta\}$ and $B = \{v \in \mathbb{R}^n : \ \|v\|_\infty \le 1\}$, contains all vectors with s nonzero entries, equal to ± 1. This is exactly the same as to say that $C_{\gamma,\beta}$ contains the convex hull of these vectors; the latter is exactly P_s. Now, γ satisfies the inclusion $P_s \subset C_{\gamma,\beta}$ if and only if for every x the support function of P_s is majorized by that of $C_{\gamma,\beta}$, namely, for every x one has

$$\begin{aligned} \max_{u\in P_s} u^T x \le \max_{y\in C(\gamma,\beta)} y^T x &= \max_{w,v} \left\{x^T A^T w + \gamma x^T v : \ \|w\|_* \le \beta, \ \|v\|_\infty \le 1\right\} \\ &= \beta\|Ax\| + \gamma\|x\|_1. \end{aligned} \tag{2.19}$$

with the convention that when $\beta = \infty$, $\beta\|Ax\|$ is ∞ or 0 depending on whether $\|Ax\| > 0$ or $\|Ax\| = 0$. That is, $P_s \subset C_{\gamma,\beta}$ if and only if

$$\max_{u\in P_s}(u^T x - \beta\|Ax\|) \le \gamma\|x\|_1.$$

By homogeneity w.r.t. x, it is equivalent to

$$\max_{u,x} \left\{u^T x - \beta\|Ax\| : u \in P_s, \|x\|_1 \le 1\right\} \le \gamma.$$

Thus, $\widehat{\gamma}_s(A)$ is the smallest γ for which the concluding inequality takes place, and we arrive at (2.17), (2.18). ∎

For $x \in \mathbb{R}^n$, let $\|x\|_{s,1}$ be the sum of the s largest magnitudes of entries in x, or, equivalently,

$$\|x\|_{s,1} = \max_{u\in P_s} u^T x.$$

Combining Theorem 2.1, Proposition 2.1 and Theorem 2.2, we get the following

Corollary 2.1 *For a matrix $A \in \mathbb{R}^{k\times n}$, $\widehat{\gamma}_s(A)$, $1 \le s \le n$ is the best upper bound of the norm $\|x\|_{s,1}$ of vectors $x \in \mathrm{Ker}(A)$ such that $\|x\|_1 \le 1$. As a result, matrix A is s-good if and only if the maximum of $\|\cdot\|_{s,1}$-norms of vectors $x \in \mathrm{Ker}(A)$ with $\|x\|_1 = 1$ is $< 1/2$.*

Note that one can treat (2.17), (2.18) as the definition of $\widehat{\gamma}_s(A,\beta)$ and then easily prove Corollary 2.1 without any reference to Theorem 2.1 and Proposition 2.1, and thus without a necessity even to introduce the characteristic $\gamma_s(A,\beta)$. However, we believe that from the methodological point of view the result of Theorem 2.1 is important, since it reveals the "true origin" of the quantities $\gamma_s(\cdot)$ and $\widehat{\gamma}_s(\cdot)$ as the entities coming from the optimality conditions for the problem (2.7).

3 Error bounds for imperfect ℓ_1-recovery via $\widehat{\gamma}$

We have seen that the quantity $\gamma_s(A)$ (or, equivalently, $\widehat{\gamma}_s(A)$) is responsible for s-goodness of a sensing matrix A, that is, for the precise ℓ_1-recovery of an s-sparse signal w in the "ideal case" when there is no measurement error and the optimization problem (2.7) is solved to exact optimality. It appears that the same quantities control the error of ℓ_1-recovery in the case when the vector $w \in \mathbb{R}^n$ is not s-sparse and the problem (2.7) is not solved to exact optimality. To see this, let w^s, $s \le n$, stand for the best, in terms of ℓ_1-norm, s-sparse approximation of w. In other words, w^s is the vector obtained from w by zeroing all coordinates except for the s largest in magnitude.

Proposition 3.1 *Let A be a $k \times n$ matrix, $1 \le s \le n$ and let $\widehat{\gamma}_s(A) < 1/2$ (or, which is the same, $\gamma_s(A) < 1$). Let also x be a ν-optimal approximate solution to the problem (2.7), meaning that*

$$Ax = Aw \ \text{ and } \|x\|_1 \le \mathrm{Opt}(Aw) + \nu,$$

where $\mathrm{Opt}(Aw)$ *is the optimal value of (2.7). Then*

$$\|x - w\|_1 \le \frac{\nu + 2\|w - w^s\|_1}{1 - 2\widehat{\gamma}_s(A)} = \frac{1 + \gamma_s(A)}{1 - \gamma_s(A)}[\nu + 2\|w - w^s\|_1].$$

Proof. Let $z = x - w$ and let I be the set of indices of s largest elements of w (i.e., the support of w^s). Denote by $x^{(s)}$ ($z^{(s)}$) the vector, obtained from x (z) by replacing by zero all coordinates of x (z) with the indices outside of I. As $Az = 0$, by Corollary 2.1,

$$\|z^{(s)}\|_1 \le \|z\|_{s,1} \le \widehat{\gamma}_s(A)\|z\|_1.$$

On the other hand, w is a feasible solution to (2.7), so $\mathrm{Opt}(Aw) \le \|w\|_1$, whence

$$\|w\|_1 + \nu \ge \|w + z\|_1 = \|w^s + z^{(s)}\|_1 + \|(w - w^s) + (z - z^{(s)})\|_1 \ge \|w^s\|_1 - \|z^{(s)}\|_1 + \|z - z^{(s)}\|_1 - \|w - w^s\|_1,$$

or, equivalently,

$$\|z - z^{(s)}\|_1 \le \|z^{(s)}\|_1 + 2\|w - w^s\|_1 + \nu.$$

Thus,

$$\begin{array}{rcl} \|z\|_1 & = & \|z^{(s)}\|_1 + \|z - z^{(s)}\|_1 \le 2\|z^{(s)}\|_1 + 2\|w - w^s\|_1 + \nu \\ & \le & 2\widehat{\gamma}_s(A)\|z\|_1 + 2\|w - w^s\|_1 + \nu, \end{array}$$

and, as $\widehat{\gamma}_s(A) < 1/2$,

$$\|z\|_1 \le \frac{2\|w - w^s\|_1 + \nu}{1 - 2\widehat{\gamma}_s(A)}.$$

∎

We study now to the properties of approximate solutions x to the problem

$$\mathrm{Opt}(y) = \min_{x \in \mathbb{R}^n} \{\|x\|_1 : \|Ax - y\| \le \varepsilon\} \tag{3.20}$$

where $\epsilon \ge 0$ and

$$y = Aw + \xi, \quad \xi \in \mathbb{R}^k,$$

with $\|\xi\| \le \varepsilon$. We are about to show that in the "non-ideal case", when w is "nearly s-sparse" and (3.20) is solved to near-optimality, the error of the ℓ_1-recovery remains "under control" – it admits an explicit upper bound governed by $\gamma_s(A, \beta)$ with a finite β. The corresponding result is as follows:

Theorem 3.1 *Let A be a $k \times n$ matrix, $s \leq n$ be a nonnegative integer, let $\epsilon \geq 0$, and let $\beta \in [0, \infty)$ be such that $\widehat{\gamma} := \widehat{\gamma}_s(A, \beta) < 1/2$. Let also $w \in \mathbb{R}^n$, let y in (3.20) be such that $\|Aw - y\| \leq \epsilon$, and let w^s be the vector obtained from w by zeroing all coordinates except for the s largest in magnitude. Assume, further, that x is a (υ, ν)-optimal solution to (3.20), meaning that*

$$\|Ax - y\| \leq \upsilon \ \ \textit{and} \ \ \|x\|_1 \leq \mathrm{Opt}(y) + \nu. \tag{3.21}$$

Then

$$\|x - w\|_1 \leq (1 - 2\widehat{\gamma})^{-1}[2\beta(\upsilon + \varepsilon) + 2\|w - w^s\|_1 + \nu]. \tag{3.22}$$

Proof. Since $\|Aw - y\| \leq \epsilon$, w is a feasible solution to (3.20) and therefore $\mathrm{Opt}(y) \leq \|w\|_1$, whence, by (3.21),

$$\|x\|_1 \leq \nu + \|w\|_1. \tag{3.23}$$

Let I be the set of indices of entries in w^s. As in the proof of Proposition 3.1 we denote by $z = x - w$ the error of the recovery, and by $x^{(s)}$ ($z^{(s)}$) the vector obtained from x (z) by replacing by zero all coordinates of x (z) with the indices outside of I. By (2.17) we have

$$\|z^{(s)}\|_1 \leq \|z\|_{s,1} \leq \beta\|Az\| + \widehat{\gamma}\|z\|_1 \leq \beta(\upsilon + \varepsilon) + \widehat{\gamma}\|z\|_1. \tag{3.24}$$

On the other hand, exactly in the same way as in the proof of Proposition 3.1 we conclude that

$$\|z\|_1 \leq 2\|z^{(s)}\|_1 + 2\|w - w^s\|_1 + \nu,$$

which combines with (3.24) to imply that

$$\|z\|_1 \leq 2\beta(\upsilon + \varepsilon) + 2\widehat{\gamma}\|z\|_1 + 2\|w - w^s\|_1 + \nu.$$

Since $\widehat{\gamma} = \widehat{\gamma}_s(A, \beta) < 1/2$, this results in

$$\|z\|_1 \leq (1 - 2\widehat{\gamma})^{-1}[2\beta(\upsilon + \varepsilon) + 2\|w - w^s\|_1 + \nu],$$

which is (3.22). ∎

The bound (3.22) can be easily rewritten in terms of $\gamma_s\left(A, \frac{\beta}{1-\widehat{\gamma}}\right) = \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1$ instead of $\widehat{\gamma} = \widehat{\gamma}_s(A, \beta)$.

The error bound (3.22) for imperfect ℓ_1-recovery, while being in some respects weaker than the RI-based bound (1.5), is of the same structure as the latter bound: assuming $\beta < \infty$ and $\widehat{\gamma}_s(A, \beta) < 1/2$ (or, equivalently, $\gamma_s(A, 2\beta) < 1$), the error of imperfect ℓ_1-recovery can be bounded in terms of $\widehat{\gamma}_s(A, \beta)$, β, measurement error ϵ, "s-tail" $\|w^s - w\|_1$ of the signal to be recovered and the inaccuracy (υ, ν) to which the estimate solves the program (3.20). The only flaw in this interpretation is that we need $\widehat{\gamma}_s(A, \beta) < 1/2$, while the "true" necessary and sufficient condition for s-goodness is $\widehat{\gamma}_s(A) < 1/2$. As we know, $\widehat{\gamma}_s(A, \beta) = \widehat{\gamma}_s(A)$ for all finite "large enough" values of β, but we do not want the "large enough" values of β to be really large, since the larger β is, the worse is the error bound (3.22). Thus, we arrive at the question "what is large enough" in our context. Here are two simple results in this direction.

Proposition 3.2 *Let A be a $k \times n$ sensing matrix of rank k.*
(i) *Let $\|\cdot\| = \|\cdot\|_2$. Then for every nonsingular $k \times k$ submatrix $\bar{A}$ of A and every $s \le k$ one has*

$$\beta \ge \bar{\beta} = \sigma^{-1}(\bar{A})\sqrt{k}, \gamma_s(A) < 1 \Rightarrow \gamma_s(A, \beta) = \gamma_s(A), \tag{3.25}$$

where $\sigma(\bar{A})$ is the minimal singular value of $\bar{A}$.
(ii) *Let $\|\cdot\| = \|\cdot\|_1$, and let for certain $\rho > 0$ the image of the unit $\|\cdot\|_1$-ball in $\mathbb{R}^n$ under the mapping $x \mapsto Ax$ contain the centered at the origin $\|\cdot\|_1$-ball of radius ρ in $\mathbb{R}^k$. Then for every $s \le k$*

$$\beta \ge \bar{\beta} = \frac{1}{\rho}, \gamma_s(A) < 1 \Rightarrow \gamma_s(A, \beta) = \gamma_s(A) \tag{3.26}$$

Proof. Given s, let $\gamma = \gamma_s(A) < 1$, so that for every vector $z \in \mathbb{R}^n$ with s nonzero entries, equal to ± 1, there exists $y \in \mathbb{R}^k$ such that $(A^T y)_i = \mathrm{sign}(x_i)$ when $x_i \ne 0$ and $|(A^T y)_i| \le \gamma$ otherwise. All we need is to prove that in the situations of (i) and (ii) we have $\|y\|_* \le \bar{\beta}$.
In the case of (i) we clearly have $\|\bar{A}^T y\|_2 \le \sqrt{k}$, whence $\|y\|_* = \|y\|_2 \le \sigma^{-1}(\bar{A})\|\bar{A}^T y\|_2 \le \sigma^{-1}(\bar{A})\sqrt{k} = \bar{\beta}$, as claimed. In the case of (ii) we have $\|A^T y\|_\infty \le 1$, whence

$$\begin{array}{l} 1 \ge \max_v \left\{v^T A^T y : v \in \mathbb{R}^n, \|v\|_1 \le 1\right\} = \max_u \left\{y^T u : u = Av, \|v\|_1 \le 1\right\} \\ \underbrace{\ge}_{(*)} \max_u \left\{u^T y : u \in \mathbb{R}^k, \|u\|_1 \le \rho\right\} = \rho\|y\|_\infty = \rho\|y\|_*, \end{array}$$

where $(*)$ is due to the inclusion $\{u \in \mathbb{R}^k : \|u\|_1 \le \rho\} \subset A\{v \in \mathbb{R}^n : \|v\|_1 \le 1\}$ assumed in (ii). The resulting inequality implies that $\|y\|_* \le 1/\rho$, as claimed. ∎

4 Efficient bounding of $\gamma_s(\cdot)$

In the previous section we have seen that the properties of a matrix A relative to ℓ_1-recovery are governed by the quantities $\widehat{\gamma}_s(A, \beta)$ – the less they are, the better. While these quantities is difficult to compute, we are about to demonstrate – and this is the primary goal of our paper – that $\widehat{\gamma}_s(A, \beta)$ admits efficiently computable "nontrivial" upper and lower bounds.

4.1 Efficient lower bounding of $\widehat{\gamma}_s(A, \beta)$

Recall that $\widehat{\gamma}_s(A, \beta) \ge \widehat{\gamma}_s(A)$ for any $\beta > 0$. Thus, in order to provide a lower bound for $\widehat{\gamma}_s(A, \beta)$ it suffices to supply such a bound for $\widehat{\gamma}_s(A)$. Theorem 2.2 suggests the following scheme for bounding $\widehat{\gamma}_s(A)$ from below. By (2.18) we have

$$\widehat{\gamma}_s(A) = \max_{u \in P_s} f(u), \ f(u) = \max_x \left\{x^T u : \|x\|_1 \le 1, Ax = 0\right\}.$$

Function $f(u)$ clearly is convex and efficiently computable: given u and solving the LP problem

$$x_u \in \mathrm{Argmax}_x \left\{u^T x : \|x\|_1 \le 1, Ax = 0\right\},$$

we get a linear form $x_u^T v$ of $v \in P_s$ which underestimates $f(v)$ everywhere and coincides with $f(v)$ when $v = u$. Therefore the easily computable quantity $\max_{v \in P_s} x_u^T v$ is a lower bound on $\widehat{\gamma}_s(A)$.

We now can use the standard *sequential convex approximation* scheme for maximizing the convex function $f(\cdot)$ over P_s. Specifically, we run the recurrence

$$u_{t+1} \in \mathrm{Argmax}_{v \in P_s} x_{u_t}^T v, \quad u_1 \in P_s,$$

thus obtaining a nondecreasing sequence of lower bounds $f(u_t) = x_{u_t}^T u_t$ on $\widehat{\gamma}_s(A)$. We can terminate this process when the improvement in bounds falls below a given threshold, and can make several runs starting from randomly chosen points u_1.

4.2 Efficient upper bounding of $\widehat{\gamma}_s(A, \beta)$.

We have seen that the representation (2.18) suggests a computationally tractable scheme for bounding $\widehat{\gamma}_s(A)$ from below. In fact, the same representation allows for a tractable way to bound $\widehat{\gamma}_s(A)$ from above, which is as follows. The difficulty in computing $\widehat{\gamma}_s(A)$ via (2.17) comes from the presence of linear equality constraints $Ax = 0$. Let us get rid of these constraints via Lagrange relaxation. Specifically, whatever be a $k \times n$ matrix Y, we clearly have

$$\max_{u,x} \left\{ u^T x : \|x\|_1 \le 1, Ax = 0, u \in P_s \right\} = \max_{u,x} \left\{ u^T (x - Y^T A x) : \|x\|_1 \le 1, Ax = 0, u \in P_s \right\},$$

whence also

$$\max_{u,x} \left\{ u^T x : \|x\|_1 \le 1, Ax = 0, u \in P_s \right\} \le \max_{u,x} \left\{ u^T (x - Y^T A x) : \|x\|_1 \le 1, u \in P_s \right\}.$$

The right hand side in this relation is easily computable, since the objective in the right hand side problem is linear in x, and the domain of x in this problem is the convex hull of just $2n$ points $\pm e_i$, $1 \le i \le n$, where e_i are the basic orths:

$$\begin{aligned} \max_{u,x} \left\{ u^T (x - Y^T A x) : \ \|x\| \le 1, u \in P_s \right\} &= \max_{\substack{u,i, \\ 1 \le i \le n}} \left\{ |u^T (I - Y^T A) e_i| : \ u \in P_s \right\} \\ &= \max_{1 \le i \le n} \max_{u \in P_s} |u^T (I - Y^T A) e_i| = \max_i \|(I - Y^T A) e_i\|_{s,1}. \end{aligned}$$

Thus, for all $Y \in \mathbb{R}^{k \times n}$,

$$\begin{aligned} \widehat{\gamma}_s(A) &= \max_{u,x} \left\{ u^T x : \|x\|_1 \le 1, Ax = 0, u \in P_s \right\} \\ &\le f_{A,s}(Y) := \max_{u \in P_s} u^T [(I - Y^T A) e_i] = \max_i \|(I - Y^T A) e_i\|_{s,1}, \end{aligned}$$

so that when setting $\alpha_s(A, \infty) := \min_Y f_{A,s}(Y)$, we get

$$\widehat{\gamma}_s(A) \le \alpha_s(A, \infty).$$

Since $f_{A,s}(Y)$ is an easy-to-compute convex function of Y, the quantity $\alpha_s(A, \infty)$ also is easy to compute (in fact, this is the optimal value in an explicit LP program with sizes polynomial in k, n).

This approach can be easily modified to provide an upper bound for $\widehat{\gamma}_s(A, \beta)$. Namely, given a $k \times n$ matrix A and $s \le k$, $\beta \in [0, \infty]$, let us set

$$\alpha_s(A, \beta) = \min_{Y = [y_1, \ldots, y_n] \in \mathbb{R}^{k \times n}} \left\{ \max_{1 \le j \le n} \|(I - Y^T A) e_j\|_{s,1} : \|y_i\|_* \le \beta, 1 \le i \le n \right\}. \tag{4.27}$$

As with γ_s, $\widehat{\gamma}_s$ we shorten the notation $\alpha_s(A,\infty)$ to $\alpha_s(A)$.

It is easily seen that the optimization problem is (4.27) is solvable, and that $\alpha_s(A,\beta)$ is nondecreasing in s, convex and nonincreasing in β, and is such that $\alpha_s(A,\beta) = \alpha_s(A)$ for all large enough values of β (cf. similar properties of $\widehat{\gamma}_s(A,\beta)$). The central observation in our context is that $\alpha_s(A,\beta)$ *is an efficiently computable upper bound on* $\widehat{\gamma}_s(A,s\beta)$, provided that the norm $\|\cdot\|$ is efficiently computable. Indeed, the efficient computability of $\alpha_s(A,\beta)$ stems from the fact that this is the optimal value in an explicit convex optimization problem with efficiently computable objective and constraints. The fact that α_s is an upper bound on $\widehat{\gamma}_s$ is stated by the following

Theorem 4.1 *One has* $\widehat{\gamma}_s(A,s\beta) \le \alpha_s(A,\beta)$.

Proof. Let I be a subset of $\{1,...,n\}$ of cardinality $\le s$, $z \in \mathbb{R}^n$ be a s-sparse vector with nonzero entries equal to ± 1, and let I be the support of z. Let $Y = [y_1,...,y_n]$ be such that $\|y_i\|_* \le \beta$ and the columns in $\Delta = I - Y^T A$ are of the $\|\cdot\|_{s,1}$-norm not exceeding $\alpha_s(A,\beta)$. Setting $y = Yz$, we have $\|y\|_* \le \beta\|z\|_1 \le \beta s$ due to $\|y_j\|_* \le \beta$ for all j. Besides this,

$$\|z - A^T y\|_\infty = \|(I - A^T Y)z\|_\infty = \|\Delta^T z\|_\infty \le \alpha_s(A,\beta),$$

since the $\|\cdot\|_{s,1}$-norms of rows in Δ^T do not exceed $\alpha_s(A,\beta)$ and z is an s-sparse vector with nonzero entries ± 1. We conclude that $\widehat{\gamma}_s(A,s\beta) \le \alpha_s(A,\beta)$, as claimed. ∎

Some comments are in order.

A. By the same reasons as in the previous section, it is important to know how large should be β in order to have $\alpha_s(A,\beta) = \alpha_s(A)$. Possible answers are as follows. Let A be a $k \times n$ matrix of rank k. Then

(i) Let $\|\cdot\| = \|\cdot\|_2$. Then for every nonsingular $k \times k$ submatrix $\bar{A}$ of A and every $s \le k$ one has

$$\beta \ge \bar{\beta} = \frac{3}{2}\sigma^{-1}(\bar{A})\sqrt{k}, \alpha_s(A) < 1/2 \Rightarrow \alpha_s(A,\beta) = \alpha_s(A), \tag{4.28}$$

where $\sigma(\bar{A})$ is the minimal singular value of $\bar{A}$.

(ii) Let $\|\cdot\| = \|\cdot\|_1$, and let for certain $\rho > 0$ the image of the unit $\|\cdot\|_1$-ball in $\mathbb{R}^n$ under the mapping $x \mapsto Ax$ contain the centered at the origin $\|\cdot\|_1$-ball of radius ρ in $\mathbb{R}^k$. Then for every $s \le k$

$$\beta \ge \bar{\beta} = \frac{3}{2\rho}, \alpha_s(A) < 1/2 \Rightarrow \alpha_s(A,\beta) = \alpha_s(A) \tag{4.29}$$

The proof is completely similar to the one of Proposition 3.2.

Note that the above bounds on β "large enough to ensure $\alpha_s(A,\beta) = \alpha_s(A)$", same as their counterparts in Proposition 3.2, whatever conservative they might be, are "constructive": to use the bound (4.28), it suffices to find a (whatever) nonsingular $k \times k$ submatrix of A and to compute its minimal singular value. To use (4.29), it suffices to solve k LP programs

$$\rho_i = \max_{x,\rho}\left\{\rho : \|x\|_1 \le 1, (Ax)_j = \rho\delta_i^j, 1 \le j \le k\right\}, \ i = 1,...,k$$

(δ_i^j are the Kronecker symbols) and to set $\rho = \min_i \rho_i$.

B. Whenever s, t are positive integers, we clearly have $\|z\|_{st,1} \le s\|z\|_{t,1}$, whence

$$\alpha_s(A,\beta) \le s\alpha_1(A,\beta). \tag{4.30}$$

Thus, we can replace in Theorem 4.1 the quantity $\alpha_s(A,\beta)$ with $s\alpha_1(A,\beta)$. As a compensation for increased conservatism, note that while both α_s and α_1 are efficiently computable, the second quantity is computationally "much cheaper". Indeed, computing $\alpha_1(A,\beta)$ reduces to solving n convex programs

$$\alpha^i := \min_{y_i} \left\{ \|e_i - A^T y_i\|_\infty : \|y_i\|_* \le \beta \right\}, \; i = 1, ..., n,$$

of design dimension k each (indeed, we clearly have $\alpha_1(A,\beta) = \max_i \alpha^i$, and the matrix Y with columns coming from optimal solutions to the outlined programs is an optimal solution to the program (4.27) corresponding to $s = 1$). In contrast to this, solving (4.27) with $s \ge 2$ seemingly cannot be decomposed in the aforementioned manner, while "as it is" (4.27) is a convex program of the design dimension kn. Unless k, n are small, solving a single optimization program of design dimension kn usually is much more demanding computationally than solving n programs of similar structure with design dimension k each.

C. Remarks in **B** point at some simple, although instructive conclusions. Let A be a $k \times n$ matrix with nonzero columns A_j, $1 \le j \le n$, and let $\mu(A)$ be its mutual incoherence, as defined in (1.6).[1)]

Proposition 4.1 *For $\beta(A) = \max\limits_{1\le j\le n} \frac{\|A_j\|_*}{\|A_j\|_2^2}$ we have*

$$\alpha_1\left(A, \frac{\beta(A)}{1+\mu(A)}\right) \le \frac{\mu(A)}{1+\mu(A)}. \tag{4.31}$$

In particular, when $\mu(A) < 1$ and $1 \le s < \frac{1+\mu(A)}{2\mu(A)}$, one has

$$\gamma_s(A, 2s\beta(A)) \le \gamma_s\left(A, \frac{s\beta(A)(1+\mu(A))}{1-(s-1)\mu(A)}\right) \le \frac{s\mu(A)}{1-(s-1)\mu(A)} < 1. \tag{4.32}$$

Proof. Indeed, with $Y_* = [A_1/\|A_1\|_2^2, ..., A_n/\|A_n\|_2^2]$, the diagonal entries in $Y_*^T A$ are equal to 1, while the off-diagonal entries are in absolute values $\le \mu(A)$; besides this, the $\|\cdot\|_*$-norms of the columns of Y_* do not exceed $\beta(A)$. Consequently, for $Y_+ = \frac{1}{1+\mu(A)} Y_*$, the absolute values of all entries in $I - Y_+^T A$ are $\le \frac{\mu(A)}{1+\mu(A)}$, while the $\|\cdot\|_*$-norms of columns of Y_+ do not exceed $\frac{\beta(A)}{1+\mu(A)}$. We see that the right hand side in the relation

$$\alpha_1\left(A, \frac{\beta(A)}{1+\mu(A)}\right) = \min_{Y=[y_1,...,y_n]} \left\{ \max_{i,j} |(I - Y^T A)_{ij}| : \|y_i\|_* \le \frac{\beta(A)}{1+\mu(A)} \right\}$$

does not exceed $\frac{\mu(A)}{1+\mu(A)}$, since Y_+ is a feasible solution for the optimization program in right hand side. This implies the bound (4.31).

[1)] Note that the "Euclidean origin" of the mutual incoherence is not essential in the following derivation. We could start with an arbitrary say, differentiable outside of the origin, norm $p(\cdot)$ on $\mathbb{R}^k$, define $\mu(A)$ as $\max\limits_{i\ne j} |A_j^T p'(A_i)|/p(A_i)$ and define $\beta(A)$ as $\max\limits_i \|p'(A_i)/p(A_i)\|_*$, arriving at the same results.

To show (4.32) note that from (4.31) with $\beta = \frac{\beta(A)}{1+\mu(A)}$ and (4.30) we have

$$\alpha_s(A,\beta) \le s\alpha_1(A,\beta) \le \frac{s\mu(A)}{1+\mu(A)},$$

and it remains to invoke Theorem 4.1 and Proposition 2.1. ∎

Observe that taken along with Theorem 3.1, bound (4.32) recovers some of the results from [13].

4.3 Application to weighted ℓ_1-recovery

Note that ℓ_1-recovery "as it is" makes sense only when A is properly normalized, so that, speaking informally, Ax is "affected equally" by all entries in x. In a general case, one could prefer to use a "weighted" ℓ_1-recovery

$$\tilde{x}_{\Lambda,\epsilon}(y) \in \mathrm{Argmin}_{z\in\mathbb{R}^n} \left\{\|\Lambda z\|_1 : \|Az - y\| \le \epsilon\right\}, \tag{4.33}$$

where Λ is a diagonal matrix with positive diagonal entries λ_i, $1 \le i \le n$, which, without loss of generality, we always assume to be ≤ 1. By change of variables $x = \Lambda^{-1}\xi$, investigating Λ-weighted ℓ_1-recovery reduces to investigating the standard recovery with the matrix $A\Lambda^{-1}$ in the role of A, followed by simple "translation" of the results into the language of the original variables. For example, the "weighted" version of our basic Theorem 3.1 reads as follows:

Theorem 4.2 *Let A be a $k \times n$ matrix, Λ be a $n \times n$ diagonal matrix with positive entries, $s \le n$ be a nonnegative integer, and let $\beta \in [0,\infty)$ be such that $\widehat{\gamma} := \widehat{\gamma}_s(A\Lambda^{-1},\beta) < 1$. Let also $w \in \mathbb{R}^n$, $\omega = \Lambda w$, and let ω^s be the vector obtained from ω by zeroing all coordinates except for the s largest in magnitude. Assume, further, that y in (4.33) satisfies the relation $\|Aw - y\| \le \epsilon$, and that x is a (υ,ν)-optimal solution to (4.33), meaning that*

$$\|Ax - y\| \le \upsilon \quad \textit{and} \quad \|\Lambda x\|_1 \le \nu + \mathrm{Opt}(y),$$

where $\mathrm{Opt}(y)$ is the optimal value of (4.33). Then

$$\|\Lambda(x - w)\|_1 \le (1 - 2\widehat{\gamma})^{-1}[2\beta(\upsilon + \varepsilon) + 2\|\omega - \omega^s\|_1 + \nu]. \tag{4.34}$$

The issue we want to address here is how to choose the scaling matrix Λ. When our goal is to recover well signals with as much nonzero entries as possible, we would prefer to make $\widehat{\gamma}_s(A\Lambda^{-1}) < 1/2$ for as large s as possible (see Theorem 2.1 and Proposition 2.1), imposing a reasonable lower bound on the diagonal entries in Λ (the latter allows to keep the left hand side in (4.34) meaningful in terms of the original variables). The difficulty is that $\widehat{\gamma}_s(A\Lambda^{-1},\beta)$ is hard to compute, not speaking about minimizing it in Λ. However, we can minimize in Λ the efficiently computable quantity $\alpha_s(A\Lambda^{-1},\bar{\beta})$, $\bar{\beta} = \beta/s$, which is an upper bound on $\widehat{\gamma}_s(A\Lambda^{-1},\beta)$. Indeed, let

$$\mathcal{Y} = \{Y = [y_1,...,y_n] : \|y_i\|_* \le \bar{\beta},\ 1 \le i \le n\}.$$

Denoting by A_i the columns of A, we have

$$\begin{aligned}
\alpha_s(A\Lambda^{-1},\bar{\beta}) &= \min_{Y\in\mathcal{Y}} \left\{\max_{1\le i\le n} \|e_i - Y^T A_i \lambda_i^{-1}\|_{s,1}\right\} \\
&= \min_{Y\in\mathcal{Y},\,\alpha} \left\{\alpha : \|e_i - Y^T A_i \lambda_i^{-1}\|_{s,1} \le \alpha,\ 1 \le i \le n\right\} \\
&= \min_{Y\in\mathcal{Y},\,\alpha} \left\{\alpha : \|\lambda_i e_i - Y^T A_i\|_{s,1} \le \alpha\lambda_i,\ 1 \le i \le n\right\},
\end{aligned}$$

so that the problem of minimizing $\alpha_s(A\Lambda^{-1},\bar{\beta})$ in Λ under the restriction $0<\ell\le\lambda_i\le 1$ on the diagonal entries of Λ reads

$$\min_{\{\lambda_i\},\,\alpha,\,Y\in\mathcal{Y}}\left\{\alpha:\|\lambda_i e_i-Y^TA_i\|_{s,1}\le\alpha\lambda_i,\ \ell\le\lambda_i\le 1,\ 1\le i\le n\right\}. \tag{4.35}$$

The resulting problem, while not being exactly convex, reduces, by bisection in α, to a "small series" of explicit convex problems and thus is efficiently solvable. In our context, the situation is even better: basically, all we want is to impose on the would-be $\widehat{\gamma}_s$ an upper bound $\widehat{\gamma}_s(A\Lambda^{-1},\beta)\le\widehat{\gamma}$ with a given $\widehat{\gamma}<1/2$, and this reduces to solving a *single* explicit convex feasibility problem

$$\text{find } \{\lambda_i\in[\ell,1]\}_{i=1}^n \text{ and } Y\in\mathcal{Y} \text{ such that } \|\lambda_i e_i-Y^TA_i\|_{s,1}\le\widehat{\gamma}\lambda_i,\ 1\le i\le n.$$

4.4 Efficient upper bounding of $\widehat{\gamma}_s(A)$: limits of performance

The bounding mechanism based on computing $\alpha_s(\cdot,\cdot)$ allows to infer, in a computationally efficient way, that a given $k\times n$ matrix A is s-good, and "with luck" s could be reasonably large, like $O(\sqrt{k/\ln(n)})$. For example, take a realization of a random $k\times n$ matrix with independent entries taking values $\pm1/\sqrt{k}$ with probabilities 1/2. For such a matrix A, with an appropriate absolute constant $O(1)$ one clearly has $\mu(A)\le O(1)\sqrt{\ln(n)/k}$ with probability $\ge 1/2$, meaning that $\gamma_s(A,2s\beta(A))\le 1/2$ for $s\le O(1)\sqrt{k/\ln(n)}$. Note that though verifiable sufficient conditions for s-goodness based on mutual incoherence are not new, see [13], our intent is to demonstrate that our machinery does allow *sometimes* to justify s-goodness for "nontrivial" values of s, like $O(\sqrt{k/\ln(n)})$. Unfortunately, the $O(\sqrt{k})$-level of values of s is the largest which can be justified via the proposed approach, unless A is "nearly square":

Proposition 4.2 *For every $k\times n$ matrix A with $n\ge 32k$, every s, $1\le s\le n$ and every $\beta\in[0,\infty]$ one has*

$$\alpha_s(A,\beta)\ge\min\left[\frac{3s}{4(s+\sqrt{2k})},\frac{1}{2}\right]. \tag{4.36}$$

In particular, in order for $\alpha_s(A,\beta)$ to be $<1/2$ (which, according to Theorems 4.1 and 2.1, is a verifiable sufficient condition for s-goodness of A), one should have $s<2\sqrt{2k}$.

Proof. Let $\alpha:=\alpha_s(A,\beta)$; note that $\alpha\le 1$.

Observe that

$$\forall v\in\mathbb{R}^n:\|v\|_2^2\le\|v\|_{s,1}^2\max[1,\frac{n}{s^2}]. \tag{4.37}$$

Postponing for a while the proof of (4.37), let us derive (4.36) from this relation. Assume, first, that $s^2\le n$. Let $Y\in\mathbb{R}^{k\times n}$ be such that $\|[I-Y^TA]_j\|_{s,1}\le\alpha$ for all j, where $[B]_j$ is j-th column of B. Setting $Q=I-Y^TA$, we get a matrix with $\|\cdot\|_{s,1}$-norms of columns $\le\alpha$. From (4.37) it follows that the Frobenius norm of Q satisfies the relation

$$\|Q\|_F^2:=\sum_{i,j}Q_{ij}^2\le\frac{n^2\alpha^2}{s^2}.$$

Consequently,

$$\|Q^T\|_F^2\le\frac{n^2\alpha^2}{s^2},$$

whence, setting

$$H = \frac{1}{2}[Q + Q^T] = I - \frac{1}{2}[Y^T A + A^T Y],$$

we get

$$\|H\|_F^2 \le \frac{n^2\alpha^2}{s^2}$$

as well. Further, the magnitudes of the diagonal entries in Q (and thus – in Q^T and in H) are at most α, whence $\mathrm{Tr}(I-H) \ge n(1-\alpha)$. The matrix $I - H = \frac{1}{2}[Y^T A + A^T Y]$ is of the rank at most $2k$ and thus has at most $2k$ nonzero eigenvalues. As we have seen, the sum of these eigenvalues is $\ge n(1-\alpha)$, whence the sum of their squares (i.e., $\|I-H\|_F^2$) is at least $\frac{n^2(1-\alpha)^2}{2k}$. We have arrived at the relation

$$\frac{n(1-\alpha)}{\sqrt{2k}} \le \|I-H\|_F \le \|I\|_F + \|H\|_F \le \sqrt{n} + \frac{n\alpha}{s}.$$

whence

$$\alpha n\left[\frac{1}{\sqrt{2k}} + \frac{1}{s}\right] \ge \frac{n}{\sqrt{2k}} - \sqrt{n} \ge \frac{3n}{4\sqrt{2k}}$$

(the concluding inequality is due to $n \ge 32k$), and (4.36) follows. We have derived (4.36) from (4.37) when $s^2 \le n$; in the case of $s^2 > n$, let $s' = \lfloor\sqrt{n}\rfloor$, so that $s' \le s$. Applying the just outlined reasoning to s' in the role of s, we get $\alpha_{s'}(A,\beta) \ge \frac{3s'}{4(s'+\sqrt{2k})}$, and the latter quantity is $\ge 1/2$ due to $n \ge 32k$ and the origin of s'. Since $s \ge s'$, we have $\alpha_s(A,\beta) \ge \alpha_{s'}(A,\beta) \ge 1/2$, and (4.36) holds true.

It remains to prove (4.37). W.l.o.g. we can assume that $v_1 \ge v_2 \ge ... \ge v_n \ge 0$ and $\|v\|_{s,1} = 1$; let us upper bound $\|v\|_2^2$ under these conditions. Setting $v_{s+1} = \lambda$, observe that $0 \le \lambda \le \frac{1}{s}$ and that for λ fixed, we have

$$\|v\|_2^2 \le \max_{v_1,...,v_s}\left\{\sum_{i=1}^s v_s^2 : \sum_{i=1}^s v_i = 1, v_i \ge \lambda,\ 1 \le i \le s\right\} + (n-s)\lambda^2.$$

The maximum of the right hand side is achieved at an extreme point of the set $\{v \in \mathbb{R}^s : \sum_i v_i = 1, v_i \ge \lambda\}$, that is, at a point where all but one of v_i's are equal to λ, and remaining one is $1-(s-1)\lambda$. Thus,

$$\begin{aligned}\|v\|_2^2 &\le [1-(s-1)\lambda]^2 + (s-1)\lambda^2 + (n-s)\lambda^2 = 1 - 2(s-1)\lambda + (s^2-2s+n)\lambda^2\\ &\le \max_{0\le\lambda\le 1/s}[1-2(s-1)\lambda + (s^2-2s+n)\lambda^2].\end{aligned}$$

The maximum in the right hand side is achieved at an endpoint of the segment $[0, 1/s]$, i.e., is equal to $\max[1, n/s^2]$, as claimed. ∎

Discussion. Proposition 4.2 is a really bad news – it shows that our verifiable sufficient condition fails to establish s-goodness when $s > O(1)\sqrt{k}$, unless A is "nearly square". This "ultimate limit of performance" is much worse than the actual values of s for which a $k \times n$ matrix A may be s-good. Indeed, it is well known, see, e.g. [9], that a random $k \times n$ matrix with i.i.d. Gaussian or ± 1 elements is, with close to 1 probability, s-good for s as large as $O(1)k/\ln(n/k)$. This is, of course, much larger than the above limit $s \le O(\sqrt{k})$. Recall, however, that we are interested in *efficiently verifiable* sufficient condition for s-goodness, and efficient verifiability has its price. At this moment we do not know whether the "price of efficiency" can be made better than the one for the proposed approach. Note, however, that for all known deterministic provably s-good $k \times n$ matrices s is $\le O(1)\sqrt{k}$, provided $n \gg k$ [12].

5 Restricted isometry property and characterization of s-goodness

Recall that the RI property (1.4) plays the central role in the existing Compressed Sensing results, like the following one: *For properly chosen absolute constants $\delta \in (0,1)$ and integer $\kappa > 1$* (e.g., for $\delta < \sqrt{2}-1$, $\kappa = 2$, see [10, Theorem 1.1]), *a matrix possessing* RI(δ, m) *property is s-good, provided that $m \geq \kappa s$.* By Theorem 2.1 it follows that with an appropriate $\delta \in (0,1)$, the RI(δ, m)-property of A implies that $\gamma_s(A) < 1$, provided $m \geq \kappa s$. Thus, the RI property possesses important implications in terms of the characterization/verifiable sufficient conditions for s-goodness as developed above. While these implications do not contribute to the "constructive" part of our results (since the RI property is seemingly difficult to verify), they certainly contribute to better understanding of our approach and integrating it into the existing Compressed Sensing theory. In this section, we present the "explicit forms" of several of those implications.

5.1 Bounding $\widehat{\gamma}_s(A)$ for RI sensing matrices

Proposition 5.1 *Let s be a positive integer, and let A be a $k \times n$ matrix possessing the* RI$(\delta, 2s)$*-property with $0 < \delta < \sqrt{2}-1$. Then*

$$\widehat{\gamma}_s(A) \leq \frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta} \quad \textit{and} \quad \gamma_s(A) \leq \frac{\sqrt{2}\delta}{1-\delta} \tag{5.38}$$

Proof. Observe that by Lemma 2.2 of [10], for any vector $h \in \mathrm{Ker}(A)$ and any index set I of cardinality $\leq m/2$ we have under the premise of Proposition:

$$\sum_{i\in I} |h_i| \leq \rho \sum_{i \notin I} |h_i|, \quad \rho = \sqrt{2}\delta(1-\delta)^{-1}.$$

This implies that for any $h \in \mathrm{Ker}(A)$ one has $\|h\|_{s,1} \leq \rho(\|h\|_1 - \|h\|_{s,1})$, that is, $\|h\|_{s,1} \leq \frac{\rho}{1+\rho}\|h\|_1$. By Corollary 2.1 it follows that $\widehat{\gamma}_s(A) \leq \frac{\rho}{1+\rho}\,(< 1/2)$, and thus $\gamma_s(A) \leq \rho\,(< 1)$. ∎

Combining Proposition 5.1 and Theorem2.1, we arrive at a sufficient condition for s-goodness in terms of RI-property identical to the one in [10, Theorem 1.1]: *a matrix A is s-good if it possesses the* RI$(\delta, 2s)$*-property with $\delta < \sqrt{2}-1$.*

The representation (2.17) of $\widehat{\gamma}_s(A, s)$ also allows to bound the value of $\widehat{\gamma}_s(A, \beta)$ and corresponding β in the case when the Restricted Eigenvalue assumption RE(m, ρ, κ) of [4] holds true. The exact formulation of the latter assumption is as follows. Let I be an arbitrary subset of indices of cardinality s; for $x \in \mathbb{R}^n$, let x^I be the vector obtained from x by zeroing all the entries with indices outside of I. A sensing matrix A is RE(s, ρ, κ) if

$$\kappa(s,\rho) := \min_{x,I} \left\{ \frac{\|Ax\|_2}{\|x^I\|_2} : \; x \in \mathbb{R}^n,\, \rho\|x^s\|_1 \geq \|x - x^s\|_1;\; \mathrm{Card}(I) = s \right\} > 0.$$

Note that the condition $\rho\|x^s\|_1 \geq \|x - x^s\|_1$ is equivalent to $\|x^s\|_1 \geq (1+\rho)^{-1}\|x\|_1$, and $\frac{\|Ax\|_2}{\|x^s\|_2} \geq \kappa$ implies that $\|x^s\|_1 \leq \kappa^{-1}\sqrt{s}\|Ax\|_2$. Thus if the RE$(s, \rho, \kappa)$ assumption holds for A, we clearly have for any $x \in \mathbb{R}^n$

$$\|x\|_{s,1} \leq \max\left\{ \frac{\sqrt{s}\|Ax\|_2}{\kappa},\; (1+\rho)^{-1}\|x\|_1 \right\}.$$

In other words, assumption $\mathrm{RE}(s,\rho,\kappa)$ implies that

$$\widehat{\gamma}_s\left(A,\frac{\sqrt{s}}{\kappa}\right)\le(1+\rho)^{-1}.$$

5.2 "Large enough" values of β

We present here an upper bound on the value of β such that $\gamma_s(A,\beta)=\gamma_s(A)$ in the case when the matrix A possesses the RI-property (cf. Proposition 3.2):

Proposition 5.2 *Let s be a positive integer, A be a $k\times n$ matrix possessing the $\mathrm{RI}(\delta,2s)$-property with $0<\delta<\sqrt{2}-1$ and $s\le n$ and let $\|\cdot\|$ be the ℓ_2-norm. Then*

$$\widehat{\gamma}_s(A,\beta)\le\frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta}\quad\textit{for all}\quad\beta\ge\frac{\sqrt{(1+\delta)s}}{1+(\sqrt{2}-1)\delta}. \tag{5.39}$$

Proof. The derivations below are rather standard to Compressed Sensing. Let us prove that

$$\forall w\in\mathbb{R}^n:\|w\|_{s,1}\le\frac{\sqrt{(1+\delta)s}}{1+(\sqrt{2}-1)\delta}\|Aw\|_2+\frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta}\|w\|_1. \tag{5.40}$$

There is nothing to prove when $w=0$; assuming $w\neq0$, by homogeneity we can assume that $\|w\|_1=1$. Besides this, w.l.o.g. we may assume that $|w_1|\ge|w_2|\ge...\ge|w_n|$. Let us set $\alpha=\|Aw\|_2$. Let us split w into consecutive s-element "blocks" $w^0,w^1,...,w^q$, so that w^0 is obtained from w by zeroing all coordinates except for the first s of them, w^1 is obtained from w by zeroing all coordinates except of those with indices $s+1,s+2,...,2s$, and so on, with evident modification for the last block w^q. By construction we have

$$w=\sum_{j=0}^q w^j,\quad\|w_0\|_1\ge\|w^1\|_1\ge...\ge\|w^q\|_1,\ \ \|w\|_1=\sum_{j=0}^q\|w^j\|_1.$$

Further, we have due to the monotonicity of $|w_i|$ and s-sparsity of all w^j:

$$j\ge1\Rightarrow\|w^j\|_2^2\le\|w^j\|_\infty\|w^j\|_1\le s^{-1}\|w^{j-1}\|_1\|w^j\|_1\le s^{-1}\|w^{j-1}\|_1^2. \tag{5.41}$$

On the other hand, due to the RI-property of A and the fact that $\|Aw\|_2=\alpha$ we have the first inequality in the following chain:

$$\begin{aligned}\alpha\sqrt{1+\delta}\|w^0+w^1\|_2&\ge\|Aw\|_2\|A(w^0+w^1)\|_2\ge(Aw)^TA(w^0+w^1)\\&=(w^0+w^1)^TA^TA(w^0+w^1)+\sum_{j=2}^q(w^0+w^1)^TA^TAw^j\\&\ge(1-\delta)\|w^0+w^1\|_2^2-\sum_{j=2}^q\sqrt{2}\delta\|w^0+w^1\|_2\|w^j\|_2,\end{aligned} \tag{5.42}$$

where we have used the "classical" RI-based relation (see [6])

$$v^TA^TAu\le\sqrt{2}\delta\|v\|_2\|u\|_2$$

for any two vectors u, $v \in \mathbb{R}^n$ with disjoint supports and such that u is s-sparse and v is $2s$-sparse. Using (5.41) we can now continue (5.42) to get

$$\begin{aligned}(1-\delta)\|w^0+w^1\|_2^2 &\le \alpha\sqrt{1+\delta}\|w^0+w^1\|_2+\sqrt{2}\delta\|w^0+w^1\|_2\, s^{-1/2}\sum_{j=1}^{q-1}\|w^j\|_1\\ &\le \alpha\sqrt{1+\delta}\|w^0+w^1\|_2+\sqrt{2}s^{-1/2}\delta\|w^0+w^1\|_2\|w-w^0\|_1.\end{aligned}$$

Since w^0 is s-sparse, we conclude that

$$\|w^0\|_1 \le \sqrt{s}\|w^0\|_2 \le \sqrt{s}\|w^0+w^1\|_2 \le \frac{\alpha\sqrt{(1+\delta)s}}{1-\delta}+\rho\|w-w^0\|_1 = \frac{\alpha\sqrt{(1+\delta)s}}{1-\delta}+\rho(1-\|w^0\|_1) \qquad [\rho=\tfrac{\sqrt{2}\delta}{1-\delta}]$$

(recall that $\|w\|_1=1$). It follows that

$$\|w^0\|_1 \le \frac{\alpha\sqrt{(1+\delta)s}}{(1+\rho)(1-\delta)}+\frac{\rho}{1+\rho} = \frac{\alpha\sqrt{(1+\delta)s}}{1+(\sqrt{2}-1)\delta}+\frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta}.$$

Recalling that $\alpha=\|Aw\|_2$, the concluding inequality is exactly (5.40) in the case of $\|w\|_1=1$. (5.40) is proved.

Invoking (2.17), (5.40) implies that with $\|\cdot\|=\|\cdot\|_2$ and with $\beta \ge \frac{\sqrt{(1+\delta)s}}{1+(\sqrt{2}-1)\delta}$, one has $\widehat{\gamma}_s(A,\beta) \le \frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta}$. ∎

It is worth to note that when using the bounds of Proposition 5.2 on $\widehat{\gamma}_s(A,\beta)$ and the corresponding β along with Theorem 3.1, we recover the classical bounds on the accuracy of the ℓ_1-recovery as those given in [9, 10].

5.3 Limits of performance of verifiable conditions for s-goodness in the case of RI sensing matrices

It makes sense to ask how conservative is the *verifiable* sufficient condition for s-goodness "$\alpha_s(A) < 1/2$" as compared to the *difficult-to-verify* RI condition "if A is RI(δ,m), then A is s-good for $s \le O(1)m$". It turns out that this conservatism is under certain control, fully compatible with the "limits of performance" of our verifiable condition as stated in Proposition 4.2. Specifically, we are about to prove that if A is RI(δ,m), then $\alpha_s(A) < 1/2$ when $s \le O(1)\sqrt{m}$: our verifiable condition "guarantees at least square root of what actually takes place". The precise statement is as follows:

Proposition 5.3 *Let a $k\times n$ matrix A possess* RI(δ,m)*-property. Then*

$$\alpha_1(A) \le \frac{\sqrt{2}\,\delta}{(1-\delta)\sqrt{m-1}}, \tag{5.43}$$

so that

$$s < \frac{(1-\delta)\sqrt{m-1}}{2\sqrt{2}\delta} \;\Rightarrow\; \alpha_s(A) \le s\alpha_1(A) < 1/2. \tag{5.44}$$

Proof. 1^0. We start with the following simple fact (cf. Proposition 5.1):

Lemma 5.1 *Let A possess* RI(δ, m)*-property. Then*

$$\widehat{\gamma}_1(A) \le \frac{\sqrt{2}\,\delta}{(1-\delta)\sqrt{m-1}}. \tag{5.45}$$

Proof. Invoking Theorem 2.2, all we need to prove is that under the premise of Lemma for every s, $1 \le s < m$, and for every $w \in \mathrm{Ker}(A)$ we have

$$\|w\|_\infty = \|w\|_{1,1} \le \widehat{\gamma}\|w\|_1.$$

To prove this fact we use again the standard machinery related to the RI-property (cf proof of Proposition 5.2): we set $t = \lfloor m/2 \rfloor$, assume w.l.o.g. that $\|w\|_1 = 1$, $|w_1| \ge |w_2| \ge ... \ge |w_n|$ and split w into q consecutive "blocks" so that the cardinality of the "blocks" is 1, $t-1, t$, t, ... I.e. the first "block" $\omega^0 \in \mathbb{R}^n$ is the vector such that $w^0_1 = w_1$ and all other coordinates vanish, w^1 is obtained from w by zeroing all coordinates except of those with indices $2, 3, ..., t$, w^2 is obtained from w by zeroing all coordinates except of those with indices $t+1, ..., 2t$, and so on, with evident modification for the last vector w^q. Acting as in the proof of Proposition 5.2, and using the relation (see [6])

$$v^T A^T A u \le \delta \|v\|_2 \|u\|_2$$

for any two t-sparse vectors u, $v \in \mathbb{R}^n$, $t \le m/2$, with disjoint supports, we obtain

$$0 = (A(w^0 + w^1))^T A w \ge (1-\delta)\|w^0 + w^1\|_2^2 - t^{-1/2}\delta \|w^0 + w^1\|_2 (1 - |w_1|)$$

whence

$$|w_1| \le \|w^0 + w^1\|_2 \le \frac{\delta}{(1-\delta)\sqrt{t}}(1 - |w_1|) \le \frac{\delta}{(1-\delta)\sqrt{t}},$$

what is (5.45). ∎

2^0. Now we are ready to complete the proof of (5.43). We already know that $\alpha_s(A) \le s\alpha_1(A)$, so all we need is to verify (5.43). The optimization program specifying $\alpha_1(A)$ can be equivalently written down as follows:

$$\alpha_1(A) = \min_{\alpha \in \mathbb{R}, Y \in \mathbb{R}^{k \times n}} \left\{ \alpha : [e_i - Y^T A_i; \alpha] \in \mathbf{K},\ 1 \le i \le n \right\}, \tag{5.46}$$

where e_i are the basic orths in $\mathbb{R}^n$, A_i are the columns of A and $\mathbf{K}$ is the cone $\{x = [u; t] \in \mathbb{R}^n \times \mathbb{R} : t \ge \|u\|_\infty\}$; note that this cone is convex, closed, with a nonempty interior and pointed, and its dual cone is

$$\mathbf{K}_* := \{y = [v; r] \in \mathbb{R}^n \times \mathbb{R} : y^T x \ge 0\, \forall x \in \mathbf{K}\} = \{y = [v; r] : r \ge \|u\|_1\}.$$

Now, (5.46) is a strictly feasible below bounded conic program (for conic problems and conic duality, see, e.g., [2]), so that by Conic Duality Theorem its optimal value is equal to the one in its conic dual:

$$\alpha_1(A) = \max_{w_i, t_i} \Big\{ \sum_{i=1}^n e_i^T w_i : A \underbrace{\begin{bmatrix} w_1^T \\ \dots \\ w_n^T \end{bmatrix}}_{W = W[w]} = 0, \|w_i\|_1 \le t_i,\ 1 \le i \le n, \sum_i t_i = 1 \Big\} \tag{5.47}$$

and the dual problem is solvable.

In fact, (5.46) is (equivalent to) an LP program, and (5.47) is (equivalent to) the LP dual of the LP reformulation of (5.46). We prefer to skip a boring LP derivation and to use the much more transparent "conic" one. For the sake of a reader not acquainted with conic programming, here is this derivation. When building the dual, we apply to the original – primal – problem a simple mechanism for deriving lower bounds on the primal optimal value, specifically, as follows. Observe that whenever $x \in \mathbf{K}$ and $y \in \mathbf{K}^*$, we have $x^T y \geq 0$; therefore choosing somehow $y_i = [-w_i; t_i] \in \mathbf{K}_*$ (or, which is the same, choosing $w_i \in \mathbb{R}^n$ and $t_i \in \mathbb{R}$ in such a way that $\|w_i\|_1 \leq t_i$), the constraints in (5.46) imply that $t_i\alpha - w_i^T[e_i - Y^T A_i] \geq 0$ for all i and every feasible solution (Y, α) to the primal problem. It follows that the scalar inequality $\alpha \sum_i t_i + \sum_i w_i^T Y A_i \geq \sum_i w_i^T e_i$ is valid on the entire feasible domain of the primal problem. If we are lucky and the left hand side in this inequality is, as a function of α, Y, identically equal to the primal objective (which happens exactly when $\sum_i t_i = 1$ and $AW = 0$), the right hand side of this inequality is a lower bound on the primal optimal value. The dual problem is nothing but maximizing this bound in the "parameters" w_i, t_i such that $[w_i; t_i] \in \mathbf{K}_*$ and "we are lucky".

Now let w_i, t_i be an optimal solution to the dual problem, let $W = W[w]$, and let $u_1, ..., u_n$ be the columns of W; note that $u_i \in \mathrm{Ker}(A)$. By Lemma 5.1 we have

$$\widehat{\gamma}_1(A) \leq \widehat{\gamma} := \frac{\sqrt{2}\,\delta}{(1-\delta)\sqrt{m-1}}. \tag{5.48}$$

Invoking Corollary 2.1 with $s = 1$, we get

$$\|u_i\|_\infty = \|u_i\|_{1,1} \leq \widehat{\gamma}\|u_i\|_1.$$

We now have

$$s\alpha_1(A) = s\sum_i w_i^T e_i = s\mathrm{Tr}(W) \leq s\sum_i \|u_i\|_\infty \leq s\widehat{\gamma}\sum_i \|u_i\|_1 = s\widehat{\gamma}\sum_i \|w_i\|_1 \leq s\widehat{\gamma}\sum_i t_i = s\widehat{\gamma}.$$

Taking into account (5.48), we arrive at (5.43). ∎

6 Numerical illustration

We are about to present some very preliminary numerical results for relatively small sensing matrices. While in no sense being conclusive, these experiments exhibit some curious phenomena.

The data. We deal with two 240×256 sensing matrices. The entries of the matrix A_{rand} were drawn at random, independently of each other, from the uniform distribution on $\{-1, 1\}$. The second matrix is constructed as follows. Let us consider a signal x "living" on $\mathbf{Z}^2$ and supported on the 16×16 grid $\Gamma = \{(i, j) \in \mathbf{Z}^2 : \ 0 \leq i, j \leq 15\}$. We subject such a signal to discrete time convolution with a kernel supported on the set $\{(i, j) \in \mathbf{Z}^2 : -7 \leq i, j \leq 7\}$, and then restrict the result on the 16×15 grid $\Gamma_+ = \{(i, j) \in \Gamma : 1 \leq j \leq 15\}$. This way we obtain a linear mapping $x \mapsto A_{\mathrm{conv}}x : \mathbb{R}^{256} \to \mathbb{R}^{240}$. The matrix A_{conv} of this mapping is our second sensing matrix.

The goal of the experiment is to bound from below and from above the maximal s for which the matrix in question is s-good (the quantity $s_*(A)$ from Definition 2.1).

Sensing matrix	lower bounds on $s_*(A)$			upper bounds on $s_*(A)$	
	$\mu(A)$-bound	α_1-bound	$\bar{\alpha}_1$-bound	SCA	Simulation
$A_{\rm rand}$	2	17	19	45	$\geq$92
$A_{\rm conv}$	0	3	4	5	$\geq$30

Table 1: Efficiently computable bounds on $s_*(A)$.

Lower bounds: $\mu(A)$-bound: the bound (4.32) based on mutual incoherence; α_1-bound: the bound based on computing $\alpha_1(A)$ and upper bounding of $\alpha_s(A)$ by $s\alpha_1(A)$; $\bar{\alpha}_1$-bound: "improved" bound based on upper bounding of $\alpha_s(A)$ via the matrix Y obtained when computing $\alpha_1(A)$. **Upper bounds:** SCA: bound based on successive convex approximation; Simulation: simulation-based bound, 1,000-element sample.

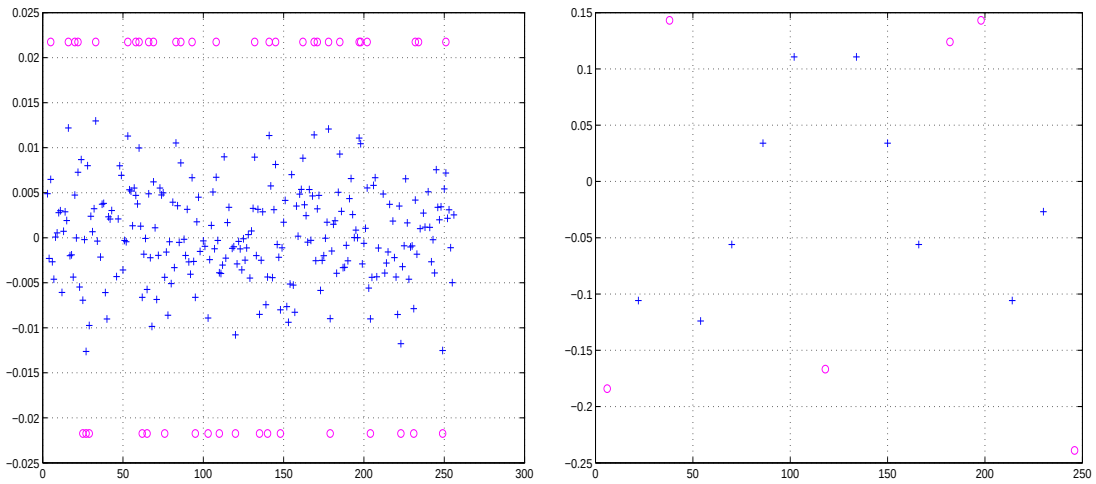
The lower bound on $s_*(A)$ was obtained via bounding from above, for various s, the quantity $\alpha_s(A)$ and invoking Theorem 4.1 and Proposition 2.1 which, taken together, state that a sufficient condition for A to be s-good is $\alpha_s(A) < 1/2$. While $\alpha_s(A)$ is efficiently computable via LP (when $\beta = \infty$, the optimization program in (4.27) is easily convertible into a linear programming one), the sizes of the resulting LP are rather large – when A is $k \times m$, the LP reformulation of (4.27) has a $(2n^2+n) \times (n(m+n+1)+1)$ constraint matrix. With our $k = 240$ and $n = 256$, the sizes of the LP become 131,328×127,233, and we preferred to avoid solving this, by no means small, LP program. Instead, we used the upper bound $\alpha_s(A) \leq s\alpha_1(A)$ (see Comment **B** in Section 4) which allows to reduce everything to computing just $\alpha_1(A)$. As explained in Comment **B**, this computation reduces to solving n convex programs of design dimension k each, and these programs are easily convertible to LP's with $(2n+1) \times (k+1)$ constraint matrices. Thus, in our experiments computing $\alpha_1(A)$ reduces to solving 256 LP's of the size 513×241. These LP's were solved by the commercial LP solver `mosekopt`, the entire computation taking less than 10 min. Note that in fact computing $\alpha_1(A)$ allows to somehow improve the trivial upper bound $s\alpha_1(A)$ on $\alpha_s(A)$, specifically, as follows. As a result of computing $\alpha_1(A)$, we get the associated matrix Y; the largest of $\|\cdot\|_{s,1}$-norms of the columns of $I - Y^TA$ clearly is an upper bound on $\alpha_s(A)$, and this bound is *at worst* $s\alpha_1(A)$.

After (upper bound on) $\alpha_s(A)$ is computed and turns out to be $< 1/2$, we know that A is s-good, and our lower bound on $s_*(A)$ is the largest s for which this nice situation takes place; note that computing this bound reduces to a *single* computation of $\alpha_1(A)$.

The lower bound on $\gamma_s(A)$ was computed using to the Sequential Convex Approximation (SCA) algorithm presented in Section 4.1. For comparison, we used also a "brute force" simulation-based approach to upper bounding of $s_*(A)$. With this approach, we generate, given s, a sample of s-sparse test vectors w and solve the associated problems (2.7) (recall that we are estimating $s_*(A)$, i.e., deal with the case of noiseless observations). We then check whether the discrepancy between the resulting optimal solution and w is well above the one which could be explained by inaccuracies of numerical ℓ_1-minimization. When this is the case, we conclude that our trial value of s is $> s_*(A)$.

The results of our experiments are presented in Table 1. On Fig. 1 we present the certificates for the upper bounds $\overline{u}$ on $s_*(A)$ given by the SCA algorithm as applied to our test matrices (see Table 1); these certificates are "poorly recovered" $(\overline{u}+1)$-sparse signals computed by the SCA. The comments are as follows:

1. Our efficiently computable lower and upper bounds on $s_*(A)$ "somehow" work in the case of the randomly chosen sensing matrix and work quite well in the case of the convolution matrix. While the gap between the lower and the upper bound in the case of the random sensing matrix



$A_{\rm rand}$, 46-sparse signal $A_{\rm conv}$, 6-sparse signal

Figure 1: Poorly recovered signals. o: true signal; +: ℓ_1-recovery, no observation noise.

could be better, we can re-iterate at this point our remark that computability has its price. We may add to this that though a random sensing matrix is a nice mathematical entity, it seems unclear how such a matrix can occur in applications – how can we measure a "completely unstructured" linear forms of a signal, or its randomly chosen Fourier coefficients, without getting access to every individual entry in the signal?

2. We see that out lower bounds on $s_*(A)$ outperform significantly those based on mutual incoherence.

3. Another intriguing phenomenon, which consistently took place in all experiments we have carried out so far, is how poor the performance of the "brute force" upper bounding is. It seems quite surprising that in 1,000 experiments with randomly chosen patterns of locations and signs of nonzero entries in an s-sparse signal with $s \geq 6s_*(A)$ (for the convolution experiment) we did not meet a *single* realization of a signal for which the inaccuracy of ℓ_1-recovery goes beyond the errors of numerical ℓ_1-minimization carried out by a good LP solver? Should we conclude that the price to be paid for the desire to recover well *all* s-sparse signals instead of a "close to 1 fraction" of them is by far too high?

References

[1] d'Aspremont, A., El Ghaoui, L., Testing the Nullspace Property using Semidefinite Programming. `http://arxiv.org/PS_cache/arxiv/pdf/0807/0807.3520v1.pdf`

[2] Ben-Tal, A., Nemirovski, A., *Lectures on Modern Convex Optimization*, SIAM, Philadelphia, 2001.

[3] Bickel, P.J. Discussion of The Dantzig selector: statistical estimation when p is much larger than n, by E.J. Candes and T. Tao. *Annals of Stat.* **35**, 23522357 (2007).

[4] Bickel P.J., Ritov, Ya., Tsybakov, A. Simultaneous analysis of Lasso and Dantzig selector, *Annals of Stat.*, to appear (2008).

[5] Candès E.J., Tao. T. The Dantzig selector: statistical estimation when p is much larger than n. *Annals of Stat.*, **35** 23132351, (2007).

[6] Candès, E.J., Tao, T. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51 , 4203-4215, (2006).

[7] Candès, E.J., Tao, T. Near-optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inform. Theory*, 52 , 5406-5425, (2006).

[8] Candès, E.J., Romberg, J., Tao, T. Signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* 59:8, 1207-1223, (2005).

[9] Candès, E.J. Compressive sampling. Marta Sanz-Solé, Javier Soria, Juan Luis Varona, Joan Verdera, Eds. *International Congress of Mathematicians, Madrid 2006*, Vol. III, 1437–1452. European Mathematical Society Publishing House, (2006).

[10] Candès, E. J. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus de l'Acad. des Sci.*, Serie I, 346, 589592 (2008).

[11] Cohen, A., Dahmen, W., DeVore, R. (2006), Compressed sensing and best k-term approximation. Submitted for publication, 2006.
E-print: `http://www.math.sc.edu/~devore/publications/CDDSensing_6.pdf`

[12] DeVore, R. Deterministic Constructions of Compressed Sensing Matrices. Preprint, Department of Mathematics, University of South Carolina, (2007).

[13] Donoho, D., Elad, M., Temlyakov V.N. Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory*, **52**, 6-18 (2006).

[14] Fuchs, J.-J. On sparse representations in arbitrary redundant bases, *IEEE Trans. Inf. Theory*, **50**, 1341-1344 (2004).

[15] Fuchs, J.-J. Recovery of exact sparse representations in the presence of bounded noise, *IEEE Trans. Inf. Theory*, **51**, 3601-3608 (2005).

[16] Meinshausen, N. and Yu, B. Lasso type recovery of sparse representations for high dimensional data, *Annals of Stat.*, to appear (2008).

[17] Tibshirani, R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B. **58**, 267-288 (1996).

[18] Tropp, J.A. Just relax: Convex programming methods for identifying sparse signals, *IEEE Trans. Info. Theory*, **51**, 3, 1030-1051 (2006).