HAL
open science

On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization

Anatoli B. Juditsky, Arkadii S. Nemirovski

To cite this version:

Anatoli B. Juditsky, Arkadii S. Nemirovski. On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization. Mathematical Programming, 2011, 127 (1), pp.57-88. 10.1007/s10107-010-0417-z . hal-00321775v2

HAL Id: hal-00321775

https://hal.science/hal-00321775v2

Submitted on 19 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization

Anatoli Iouditski* Arkadi Nemirovski†

April 19, 2010

Abstract

We discuss 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 [4-10,18-22] and references therein) the emphasis is on assessing sparse signal $w \in \mathbb{R}^n$ from an observation $y \in \mathbb{R}^k$ (in this context $k \ll n$):

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

where $\|\cdot\|$ is a given norm on $\mathbb{R}^k$, ξ is the observation error and $\varepsilon \geq 0$ 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

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

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

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

and Lasso estimator, see [21, 4], 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 w is close to w, provided the observation error ϵ is small. For instance, necessary and sufficient conditions of exactness of ℓ_1-recovery

*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 # N000140811104.

in the case of noiseless observation (when $\varepsilon = 0$) has been established in [23, 16, 15]. Specifically, in [23] it is shown that w is the unique solution of the noiseless ℓ_1-recovery problem

$$\min_z \{\|z\|_1 : \ Az = Aw\} . \tag{1.4}$$

if and only if the kernel Ker$A$ of the sensing matrix is *strict s-balanced*, the latter meaning that for any set $I \subset \{1, ..., n\}$ of cardinality $\leq s$ it holds

$$\sum_{i \in I} |z_i| < \sum_{i \notin I} |z_i| \ \text{ for any } \ z \in \mathrm{Ker}A \tag{1.5}$$

(what the above condition is sufficient for the ℓ_1-recovery to be exact in the noiseless case was stated in [14]).

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

$$\sqrt{1-\delta}\|x\|_2 \leq \|Ax\|_2 \leq \sqrt{1+\delta}\|x\|_2 \text{ for all } x \in \mathbb{R}^n \text{ with at most } m \text{ nonzero entries.} \tag{1.6}$$

For instance, the following result is well known ([10, Theorem 1.2] or [9, Theorem 4.1]): let $\|\cdot\|$ in (1.1) be the Euclidean norm $\|\cdot\|_2$, and let the sensing matrix A satisfy RI$(\delta, 2s)$-property with $\delta < \sqrt{2} - 1$. Then

$$\|\widehat{w} - w\|_1 \quad \leq \quad 2(1-\rho)^{-1}[\alpha\varepsilon\sqrt{s} + (1+\rho)\|w - w^s\|_1] \tag{1.7}$$

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$(1/4, 2s)$-property for "large" s – as large as $O\left(k/\ln(n/k)\right)$. 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$(\delta, 2s)$- property for such large values of s. For all known deterministic families of $k \times n$ matrices provably possessing the RI$(\delta, 2s)$-property, one has $s = O(\sqrt{k})$ (see [13]), which is essentially worse than the bound $s = O(1)\left(k/\ln(n/k)\right)$ promised by the RI-based theory. Moreover, RI-property itself is "intractable" – the only currently available technique to verify the RI(δ, m) 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 (1.7), 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 strict s-balancedness, nor Restricted Isometry, or Restricted Correlation assumption and the like, do 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 2- or 3-dimensional n-point grid E of possible locations of signal sources and an N-element grid R of possible locations of sensors. A sensor at a given

location measures a known linear form of the signals emitted at the nodes of E which depends on location, 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*

$$\mu(A) = \max_{i \neq j} \frac{|A_i^T A_j|}{A_i^T A_i} \tag{1.8}$$

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 (1.7) can be obtained for matrices with small mutual incoherence: a matrix A with mutual incoherence $\mu(A)$ and columns A_j of unit $\|\cdot\|_2$-norm satisfies $\mathrm{RI}(\delta, m)$ assumption (1.6) with $\delta = (m-1)\mu(A)$. 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.

The goal of this paper is to provide new computationally tractable sufficient conditions for sparse recovery.

The overview of our main results is as follows.

1. Let for $x \in \mathbb{R}^n$

$$\|x\|_{s,1} = \max_{\mathrm{Card}(I) \leq s} \sum_{i \in I} |x_i|$$

stand for the sum of s maximal magnitudes of components of x. Set

$$\widehat{\gamma}_s(A) = \max_x \{\|x\|_{s,1} : \ \|x\|_1 \leq 1, \ Ax = 0\}.$$

Starting from optimality conditions for the problem (1.4) of noiseless ℓ_1-recovery, we show that A is s-good if and only if $\widehat{\gamma}_s(A) < 1/2$, thus recovering some of the results of [23]. While $\widehat{\gamma}_s(A)$ 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, in order to cope with the case of imperfect ℓ_1-recovery (nonzero observation error, nearly s-sparse true signal, (1.2) is not solved to exact optimality), we embed the characteristic $\widehat{\gamma}_s(A)$ into a single-parametric family of characteristics $\widehat{\gamma}_s(A, \beta)$, $0 \leq \beta \leq \infty$. Here

$$\widehat{\gamma}_s(A, \beta) = \max_x \{\|x\|_{s,1} - \beta\|Ax\| : \ \|x\|_1 \leq 1\}$$

(note 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 $\widehat{\omega}$ admits an explicit upper bound, similar in structure the RI-based bound (1.7):

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

where ε is the measurement error and ν is the inaccuracy in solving (1.2).

2. The characteristics $\widehat{\gamma}_s(A, \beta)$ is still difficult to compute. In Section 4, we develop efficiently computable lower and upper bounds on $\widehat{\gamma}_s(A, \beta)$. In particular, we show that the quantity $\alpha_s(A, \beta)$,

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

(here $\|\cdot\|_*$ is the norm conjugate to $\|\cdot\|$) is an upper bound on $\widehat{\gamma}_s(A, s\beta)$.

This bound provides us with an efficiently verifiable (although perhaps conservative) *sufficient condition* for s-goodness of A, namely, $\alpha_s(A,\beta) < 1/2$. We demonstrate that our verifiable sufficient conditions for s-goodness are less restrictive than those based on mutual incoherence. On the other hand, the proposed lower bounds on $\widehat{\gamma}_s(A,\beta)$ allow to bound from above the values of s for which A is s-good.

We also study limitations 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, level $O(k/\ln(n/k))$ of s for which A may be s-good, this "limit of performance" of our machinery nearly coincides with the best known values of s for which *explicitly given individual s-good $k \times n$ sensing matrices* are known.

3. 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 A is s-good for $s = O(1)m$), 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 s of goodness".

4. 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 become aware of the preprint [12] which contain results closely related to some of those in our paper. The authors of [12] have "extracted" from [11] the sufficient condition $\widehat{\gamma}_s(A) < 1/2$ for s-goodness of A and proposed an efficiently computable upper bound on $\widehat{\gamma}_s(A)$ based on semidefinite relaxation. This bound is essentially different from our, and it could be interesting to find out if one of these bounds is "stronger" than the other.

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 1 *Let A be a $k \times n$ matrix and s be an integer, $0 \le s \le n$. 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*

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = Aw\}. \tag{2.9}$$

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 $s_*(A) \le \min[k, n]$ for every $k \times n$ matrix A.

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

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

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

Definition 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*

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

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*

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

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

$$\|y\|_* \leq \beta \ \& \ (A^T y)_i \begin{cases} = z_i, & z_i \neq 0 \\ \in [-\gamma_s(A,\beta), \gamma_s(A,\beta)], & z_i = 0 \end{cases} ; \tag{2.12}$$

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

$$\|\widehat{y}\|_* \leq \beta \ \& \ \|A^T \widehat{y} - z\|_\infty \leq \widehat{\gamma}_s(A,\beta). \tag{2.13}$$

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 (2.12) 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 (2.13) 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 quantities $\gamma_s(\cdot,\cdot)$ and $\widehat{\gamma}_s(\cdot,\cdot)$ stems from the following

Theorem 1 *Let A be a $k \times n$ matrix and $s \leq n$ be a nonnegative integer.*
(i) *A is s-good if and only if $\gamma_s(A) < 1$.*
(ii) *For every $\beta \in [0,\infty]$ one has*

$$\begin{array}{ll} (a) & \gamma := \gamma_s(A,\beta) < 1 \Rightarrow \widehat{\gamma}_s\left(A, \frac{1}{1+\gamma}\beta\right) = \frac{\gamma}{1+\gamma} < 1/2; \\ (b) & \widehat{\gamma} := \widehat{\gamma}_s(A,\beta) < 1/2 \Rightarrow \gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\beta\right) = \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1. \end{array} \tag{2.14}$$

The proof of Theorem 1 is given in Appendix A.

Theorem 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 $\widehat{\gamma}_s(\cdot)$.

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

According to Theorem 1 (ii), 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:

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

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 *Consider the polytope*

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

One has

$$\widehat{\gamma}_s(A,\beta) = \max_{u,x} \left\{u^T x - \beta\|Ax\| : \ u \in P_s, \ \|x\|_1 \le 1\right\}. \tag{2.15}$$

In particular,

$$\widehat{\gamma}_s(A) = \max_{u,x} \left\{u^T x : \ u \in P_s, \ \|x\|_1 \le 1, \ Ax = 0\right\}. \tag{2.16}$$

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

$$\begin{aligned} \max_{u\in P_s} u^T x \le \max_{y \in C(\gamma,\beta)} y^T x &= \max_{w,v} \left\{x^T A^T w + \gamma x^T v : \ \|w\|_* \le \beta, \ \|v\|_\infty \le 1\right\} \\ &= \beta\|Ax\| + \gamma\|x\|_1. \end{aligned} \tag{2.17}$$

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 (2.15), (2.16). ∎

Recall that for $x \in \mathbb{R}^n$, $\|x\|_{s,1}$ is 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 1, and Theorem 2, we get the following

Corollary 1 *For a matrix $A \in \mathbb{R}^{k\times n}$ one has $\widehat{\gamma}_s(A) = \max_x\{\|x\|_{s,1} : Ax = 0, \|x\|_1 \le 1\}$, $1 \le s \le n$. 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 (2.15) and (2.16) can be seen as an equivalent definition of $\widehat{\gamma}_s(A,\beta)$, and one can easily prove Corollary 1 without any reference to Theorem 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 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 (2.9).

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 (2.9) 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 (2.9) is not solved to exact optimality. To see this, let w^s, $s \leq n$, 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 1 *Let A be a $k \times n$ matrix, $1 \leq s \leq n$ 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 (2.9), meaning that*

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

where $\mathrm{Opt}(Aw)$ is the optimal value of (2.9). Then

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

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 1,

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

On the other hand, w is a feasible solution to (2.9), so $\mathrm{Opt}(Aw) \leq \|w\|_1$, whence

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

or, equivalently,

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

Thus,

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

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

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

■

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

$$\mathrm{Opt}(y) = \min_{x \in \mathbb{R}^n} \{\|x\|_1 : \|Ax - y\| \leq \varepsilon\} \tag{3.18}$$

where $\epsilon \geq 0$ and

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

with $\|\xi\| \leq \varepsilon$. We are about to show that in the "non-ideal case", when w is "nearly s-sparse" and (3.18) 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 *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 (3.18) 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 (3.18), meaning that*

$$\|Ax - y\| \leq \upsilon \ \text{ and } \ \|x\|_1 \leq \mathrm{Opt}(y) + \nu. \tag{3.19}$$

Then

$$\|x - w\|_1 \leq (1 - 2\widehat{\gamma})^{-1}[2\beta(\upsilon + \varepsilon) + 2\|w - w^s\|_1 + \nu]. \tag{3.20}$$

Proof. Since $\|Aw - y\| \leq \epsilon$, w is a feasible solution to (3.18) and therefore $\mathrm{Opt}(y) \leq \|w\|_1$, whence, by (3.19),

$$\|x\|_1 \leq \nu + \|w\|_1. \tag{3.21}$$

Let I be the set of indices of entries in w^s. As in the proof of Proposition 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 (2.15) we have

$$\|z^{(s)}\|_1 \leq \|z\|_{s,1} \leq \beta\|Az\| + \widehat{\gamma}\|z\|_1 \leq \beta(\upsilon + \varepsilon) + \widehat{\gamma}\|z\|_1. \tag{3.22}$$

On the other hand, exactly in the same way as in the proof of Proposition 1 we conclude that

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

which combines with (3.22) 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 (3.20). ∎

The bound (3.20) 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 (3.20) for imperfect ℓ_1-recovery, while being in some respects weaker than the RI-based bound (1.7), 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 (3.18). 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 (3.20). Thus, we arrive at the question "what is large enough" in our context. Here are two simple results in this direction.

Proposition 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 $s \leq k$ one has*

$$\beta \geq \bar{\beta} = \sigma^{-1}(\bar{A})\sqrt{k}, \gamma_s(A) < 1 \Rightarrow \gamma_s(A, \beta) = \gamma_s(A), \tag{3.23}$$

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 ball $B = \{u \in \mathbb{R}^k : \|u\|_1 \leq \rho\}$. Then for every $s \leq k$*

$$\beta \geq \bar{\beta} = \frac{1}{\rho}, \gamma_s(A) < 1 \Rightarrow \gamma_s(A, \beta) = \gamma_s(A) \tag{3.24}$$

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 $x_i \neq 0$ and $|(A^T y)_i| \leq \gamma$ otherwise. All we need is to prove that in the situations of (i) and (ii) we have $\|y\|_* \leq \bar{\beta}$.
In the case of (i) we clearly have $\|\bar{A}^T y\|_2 \leq \sqrt{k}$, whence $\|y\|_* = \|y\|_2 \leq \sigma^{-1}(\bar{A})\|\bar{A}^T y\|_2 \leq \sigma^{-1}(\bar{A})\sqrt{k} = \bar{\beta}$, as claimed. In the case of (ii) we have $\|A^T y\|_\infty \leq 1$, whence

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

where $(*)$ is due to the inclusion $\{u \in \mathbb{R}^k : \|u\|_1 \leq \rho\} \subset A\{v \in \mathbb{R}^n : \|v\|_1 \leq 1\}$ assumed in (ii). The resulting inequality implies that $\|y\|_* \leq 1/\rho$, 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 $\widehat{\gamma}_s(A, \beta) \geq \widehat{\gamma}_s(A)$ 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 suggests the following scheme for bounding $\widehat{\gamma}_s(A)$ from below. By (2.16) we have

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

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

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

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 (2.16) 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. Whatever be a $k \times n$ matrix Y, we clearly have

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

whence also

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

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:

$$\begin{array}{rcl}\max_{u,x}\left\{u^T(x - Y^TAx) : \ \|x\|_1 \le 1, u \in P_s\right\} & = & \max\limits_{\substack{u,i,\\ 1\le i\le n}} \left\{|u^T(I - Y^TA)e_i| : \ u \in P_s\right\} \\ & = & \max\limits_{1\le i\le n}\max\limits_{u\in P_s} |u^T(I - Y^TA)e_i| = \max\limits_i \|(I - Y^TA)e_i\|_{s,1}.\end{array}$$

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

$$\begin{array}{rcl}\widehat{\gamma}_s(A) & = & \max\limits_{u,x}\left\{u^Tx : \|x\|_1 \le 1, Ax = 0, u \in P_s\right\} \\ & \le & f_{A,s}(Y) := \max\limits_{u\in P_s} u^T[(I - Y^TA)e_i] = \max\limits_i \|(I - Y^TA)e_i\|_{s,1},\end{array}$$

so that when setting $\alpha_s(A,\infty) := \min\limits_Y f_{A,s}(Y)$, 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

$$\alpha_s(A,\beta) = \min_{Y=[y_1,...,y_n]\in\mathbb{R}^{k\times n}} \left\{\max_{1\le j\le n} \|(I - Y^TA)e_j\|_{s,1} : \|y_i\|_* \le \beta, 1 \le i \le n\right\}. \tag{4.25}$$

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 (4.25) 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 *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 $\Delta = I - Y^TA$ 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,

$$\|z - A^Ty\|_\infty = \|(I - A^TY)z\|_\infty = \|\Delta^Tz\|_\infty \le \alpha_s(A,\beta),$$

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 $s\leq k$ one has

$$\beta\geq\bar{\beta}=\frac{3}{2}\sigma^{-1}(\bar{A})\sqrt{k},\alpha_s(A)<1/2\Rightarrow\alpha_s(A,\beta)=\alpha_s(A), \tag{4.26}$$

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 $s\leq k$

$$\beta\geq\bar{\beta}=\frac{3}{2\rho},\alpha_s(A)<1/2\Rightarrow\alpha_s(A,\beta)=\alpha_s(A) \tag{4.27}$$

The proof is completely similar to the one of Proposition 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 2, whatever conservative they might be, are "constructive": to use the bound (4.26), it suffices to find a (whatever) nonsingular $k\times k$ submatrix of A and to compute its minimal singular value. To use (4.27), it suffices to solve k LP programs

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

(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 $\|z\|_{st,1}\leq s\|z\|_{t,1}$, whence

$$\alpha_s(A,\beta)\leq s\alpha_1(A,\beta). \tag{4.28}$$

Thus, we can replace in Theorem 4 the quantity $\alpha_s(A,\beta)$ with $s\alpha_1(A,\beta)$. Further, we have $\alpha_1(A,\beta)=\max_i\alpha^i$, where

$$\alpha^i:=\min_{y_i}\left\{\|e_i-A^Ty_i\|_\infty:\ \|y_i\|_*\leq\beta\right\},\ i=1,...,n. \tag{4.29}$$

On the other hand, we have

$$\begin{aligned}\alpha^i &= \min_y\max_j\left\{[e_i-A^Ty]_j:\ \|y_i\|_*\leq\beta\right\}=\min_y\max_x\left\{(e_i-A^Ty)^Tx:\ \|y_i\|_*\leq\beta,\ \|x\|_1\leq 1\right\}\\ &= \max_x\min_y\left\{e_i^Tx-y^TAx:\ \|y_i\|_*\leq\beta,\ \|x\|_1\leq 1\right\}=\max_x\left\{e_i^Tx-\beta\|Ax\|:\ \|x\|_1\leq 1\right\}\leq\widehat{\gamma}_1(A,\beta),\end{aligned}$$

and by Theorem 4 we conclude that

$$\alpha_1(A,\beta)=\widehat{\gamma}_1(A,\beta), \tag{4.30}$$

i.e. the relaxation for $\widehat{\gamma}_1(A,\beta)$ is exact. As a compensation for increased conservatism of the bound (4.28), 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 (4.29) of design dimension k each. In contrast to this, solving (4.25) with $s\geq 2$ seemingly cannot be decomposed in the aforementioned manner, while "as it is" (4.25) 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.

4.3 Relation to the mutual incoherence condition

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 (1.8).[1)]

Proposition 3 *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

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

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).

To show (4.32) note that from (4.31) with $\beta = \frac{\beta(A)}{1+\mu(A)}$ and (4.28) 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 and Theorem 1 (ii). ∎

Observe that taken along with Theorem 3, bound (4.32) recovers some of the results from [17].

Proposition 3 implies that computing $\alpha_s(\cdot,\cdot)$ allows to infer that a given $k \times n$ matrix A is s-good, for "reasonably large" values of s. Indeed, take a realization of a random $k \times n$ matrix with independent entries taking values $\pm 1/\sqrt{k}$ 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 such verifiable sufficient conditions for s-goodness based on mutual incoherence are certainly not new, see [17]. We use them here to show that our machinery does allow *sometimes* to justify s-goodness for "nontrivial" values of s, like $O(\sqrt{k/\ln(n)})$.

4.4 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

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

[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.

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 reads as follows:

Theorem 5 *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 $\widehat{\gamma} := \widehat{\gamma}_s(A\Lambda^{-1}, \beta) < 1/2$. 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 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

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

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

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

4.5 Limits of performance

As we have seen in Section 4.3, the bounding mechanism based on computing $\alpha_s(\cdot, \cdot)$ allows to certify s-goodness of an $k \times n$-sensing matrix for s as large as $O(\sqrt{k/\ln(n)})$. 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".

4.6 $\sqrt{k}$-bound

Proposition 4 *For every $k \times n$ matrix A with $n \geq 32k$, every s, $1 \leq s \leq n$ and every $\beta \in [0, \infty]$ one has*

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

In particular, in order for $\alpha_s(A, \beta)$ to be $< 1/2$ (which, according to Theorems 4 and 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 $\alpha \leq 1$.

Observe that

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

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

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

Consequently,

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

whence, setting

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

we get

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

as well. Further, the magnitudes of the diagonal entries in Q (and thus – in Q^T and in H) are at most α, whence $\mathrm{Tr}(I - H) \geq n(1 - \alpha)$. 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 $\geq n(1 - \alpha)$, 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

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

whence

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

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

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

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

The maximum of the right hand side is achieved at an extreme point of the set $\{v \in \mathbb{R}^s : \sum_i v_i = 1, v_i \geq \lambda\}$, 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,

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

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 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 $s \leq O(\sqrt{k})$. 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 $\leq O(1)\sqrt{k}$, provided $n \gg k$ [13].

5 Restricted isometry property and characterization of s-goodness

Recall that the RI property (1.6) 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* $\mathrm{RI}(\delta, m)$ *property is s-good, provided that $m \geq \kappa s$.* By Theorem 1 it follows that with an appropriate $\delta \in (0,1)$, the $\mathrm{RI}(\delta, m)$-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 *Let s be a positive integer, and let A be a $k \times n$ matrix possessing the* $\mathrm{RI}(\delta, 2s)$*-property with $0 < \delta < \sqrt{2}-1$. Then*

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

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 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 and Theorem 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* $\mathrm{RI}(\delta, 2s)$*-property with $\delta < \sqrt{2}-1$.*

The representation (2.15) also allows to bound the value of $\widehat{\gamma}_s(A,\beta)$ and corresponding β in the case when the Restricted Eigenvalue assumption $\mathrm{RE}(m,\rho,\kappa)$ 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 $\mathrm{RE}(s,\rho,\kappa)$ 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 $\mathrm{RE}(s,\rho,\kappa)$ 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

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

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 2):

Proposition 6 *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 $s\leq n$ and let $\|\cdot\|$ be the ℓ_2-norm. Then*

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

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

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

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 $|w_1|\geq|w_2|\geq...\geq|w_n|$. 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

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

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

$$j\geq1\Rightarrow\|w^j\|_2^2\leq\|w^j\|_\infty\|w^j\|_1\leq s^{-1}\|w^{j-1}\|_1\|w^j\|_1\leq s^{-1}\|w^{j-1}\|_1^2. \tag{5.41}$$

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:

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

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

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

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

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

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

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

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

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

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

Invoking (2.15), (5.40) implies that with $\|\cdot\| = \|\cdot\|_2$ and with $\beta \geq \frac{\sqrt{(1+\delta)s}}{1+(\sqrt{2}-1)\delta}$, one has $\widehat{\gamma}_s(A,\beta) \leq \frac{\sqrt{2}\delta}{1+(\sqrt{2}-1)\delta}$. ∎

It is worth to note that when using the bounds of Proposition 6 on $\widehat{\gamma}_s(A,\beta)$ and the corresponding β along with Theorem 3, we recover the classical bounds on the accuracy of the ℓ_1-recovery as those given in [9, 10].

5.3 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 $s \leq O(1)m$". 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. Specifically, we are about to prove that if A is RI(δ, m), then $\alpha_s(A) < 1/2$ when $s \leq O(1)\sqrt{m}$: our verifiable condition "guarantees at least square root of what actually takes place". The precise statement is as follows:

Proposition 7 *Let a $k \times n$ matrix A possess* $\mathrm{RI}(\delta, m)$*-property. Then*

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

so that

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

Proof. 1^0. We start with the following simple fact (cf. Proposition 5):

Lemma 1 *Let A possess* $\mathrm{RI}(\delta, m)$*-property. Then*

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

Proof. Invoking Theorem 2, all we need to prove is that under the premise of Lemma for every s, $1 \leq s < m$, and for every $w \in \mathrm{Ker}(A)$ we have

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

To prove this fact we use again the standard machinery related to the RI-property (cf proof of Proposition 6): we set $t = \lfloor m/2 \rfloor$, assume w.l.o.g. that $\|w\|_1 = 1$, $|w_1| \geq |w_2| \geq ... \geq |w_n|$ 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 $w_1^0 = w_1$ 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 6, and using the relation (see [6])

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

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

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

whence

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

what is (5.45). ∎

2^0. Now we are ready to complete the proof of (5.43). We already know that $\alpha_s(A) \leq s\alpha_1(A)$, so all we need is to verify (5.43). The latter is readily given by (4.30) combined with (5.45). ∎

6 Numerical illustration

We are about to present some very preliminary numerical results for relatively small sensing matrices.

The data. In the two series of experiments presented below we deal with sensing matrices of row dimension $n = 256$ and $n = 1024$.

For $n = 256$ we generate three sets of random matrices of column dimension $m = 0.1n$, ..., $0.9n$: Gaussian matrices, with the i.i.d. normal entries, Fourier matrices, which are m rows of the Fourier basis on $[0, 1]$

drawn at random and, finally, Hadamard matrices, which are, again, random $m \times n$ cuts from the $n \times n$ Hadamard matrix.[2] Then all matrices are normalized so that their columns have unit ℓ_2-norm.

For $n = 1024$ we provide the result of an experiment with a family of Gaussian matrices of column dimension $m = 0.1n$, ..., $0.9n$ and with a 992×1024 matrix $A_{\rm conv}$ which is constructed as follows. Let us consider a signal x "living" on $\mathbf{Z}^2$ and supported on the 32×32 grid $\Gamma = \{(i,j) \in \mathbf{Z}^2 : \ 0 \le i,j \le 31\}$. We subject such a signal to discrete time convolution with a kernel supported on the set $\{(i,j) \in \mathbf{Z}^2 : -7 \le i,j \le 7\}$, and then restrict the result on the 32×31 grid $\Gamma_+ = \{(i,j) \in \Gamma : 1 \le j \le 31\}$. This way we obtain a linear mapping $x \mapsto A_{\rm conv}x : \ \mathbb{R}^{1024} \to \mathbb{R}^{992}$.

The goal of the experiment is to bound from below and from above the maximal s for which the $m \times n$ matrix A in question is s-good (the quantity $s_*(A)$ from Definition 1).

The lower bound on $s_*(A)$ was obtained via bounding from above, for various s, the quantity $\alpha_s(A)$ and invoking Theorem 4 and Theorem 1 (ii) which, taken together, state that a sufficient condition for A to be s-good is $\alpha_s(A) < 1/2$.

We provide two lower bounds for $s_*(A)$. The first bound is obtained using the upper bound $\alpha_s(A) \le s\alpha_1(A)$ (see Comment **B** in Section 4). When the upper bound $s\alpha_1(A)$ for $\alpha_s(A)$ 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 situation takes place; note that computing this bound reduces to a *single* computation of $\alpha_1(A)$. As explained in Comment **B**, this computation reduces to solving n convex programs of design dimension m each, and these programs are easily convertible to LP's with $(2n+1) \times (m+1)$ constraint matrices. These LP's were solved using the commercial LP solver `mosekopt` [1]. 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* $s\alpha_1(A)$.

For "small" matrices with the row dimension $n = 256$ we also provide the "improved" lower bound, obtained using the computation of $\alpha_s(A)$ itself. We act as follows: when the bound $s(\alpha_1)$ is computed, verify if the value $s(\alpha_1) + 1$ can be certified lower bound for $s_*(A)$ using the computation of $\alpha_s(A)$. If this bound is certified we proceed with $s(\alpha_1) + 2$, and so on. Note that, exactly as it is in the case of $\alpha_1(A)$, computing $\alpha_s(A)$ allows to improve the lower bound on $s_*(A)$ in the case when $\alpha_s(A) < 1/2$. Indeed, as a result of computing $\alpha_s(A)$, we get the associated matrix Y; the largest s such that the $\|\cdot\|_{s,1}$-norms of the columns of $I - Y^TA$ is $< 1/2$ clearly is a lower bound on $s_*(A)$.

We would like to add here two words about the techniques used to compute the corresponding bound (being of interest by themselves, these techniques are the subject of a separate paper). While $\alpha_s(A)$ is efficiently computable via LP (when $\beta = \infty$, the optimization program in (4.25) is easily convertible into a linear programming one), the sizes of the resulting LP are rather large – when A is $m \times n$, the LP reformulation of (4.25) has a $(2n^2+n) \times (n(m+n+1)+1)$ constraint matrix. For instance, for $m = 230$ and $n = 256$, the size of the LP becomes 131,328×127,233, and we preferred to avoid solving this, by no means small, LP program using the interior-point solver available with `mosekopt`. Instead, the LP is reformulated as a saddle-point problem and is solved using an implementation of the non-Euclidean mirror-prox algorithm, described in [20].

The upper bound on $s_*(A)$ is computed using the lower bound on $\gamma_s(A)$ by the Sequential Convex Approximation algorithm presented in Section 4.1.

The results of our experiments are presented in Tables 1 and 2. The computations we run on an Intel P9500@2.53GHz CPU (the computations were running single-core). We present along with the results the

[2] Hadamard matrix H_ℓ of order $n = 2^\ell$ is the orthogonal matrix with entries ± 1 given by the recurrence $H_0 = 1$, $H_{\ell+1} = [H_\ell, H_\ell; H_\ell, -H_\ell]$.

corresponding CPU usage.

We would like to add the following comment: our efficiently computable lower bounds on $s_*(A)$ outperform significantly those based on mutual incoherence. Further, these lower and upper bounds "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 could be better, we can re-iterate at this point our remark that computability has its price.

References

[1] Andersen, E. D., Andersen, K. D. *The MOSEK optimization tools manual. Version 5.0* http://www.mosek.com/fileadmin/products/5_0/tools/doc/html/tools/index.html

[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] Cands, 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. *Compressed Sensing and Best k-term Approximation.* Preprint, http://www.math.sc.edu/~devore/publications/CDDSensing_6.pdf (2006).

[12] d'Aspremont, A., El Ghaoui, L. *Testing the Nullspace Property using Semidefinite Programming.* Preprint, http://arxiv.org/abs/0807.3520 (2008).

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

[14] Donoho, D. Huo, X., Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory* **47**(7), 2845-2862 (2001).

[15] Donoho, D. *High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension.* Technical report, Department of Statistics, Stanford University, 2004.

[16] Donoho, D. *Neighborly polytopes and sparse solutions of underdetermined linear equations.* Technical report, Department of Statistics, Stanford University, 2004.

[17] 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).

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

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

[20] Nemirovski, A., Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM Journal on Optimization*, **15**, 229-251 (2004).

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

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

[23] Zhang, Y. *A simple proof for recoverability of ell-1-minimization: go over or under?* Rice CAAM Department Technical Report TR05-09 (2005).

A Proof of Theorem 1

Proof. (i): 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 (2.9). From the fact that w is an optimal solution to this problem it follows, by optimality conditions, that for certain y the function $f_y(x) = \|x\|_1 - y^T Ax$ attains its minimum over $x \in \mathbb{R}^n$ at $x = w$, meaning that $0 \in \partial f_y(w)$, that is,

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

so that the LP problem

$$\min_{y,\gamma} \left\{ \gamma : (A^T y)_i \begin{cases} = \text{sign}(w_i), & i \in I \\ \in [-\gamma, \gamma], & i \in \bar{I} \end{cases} \right\} \tag{A.46}$$

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 $\{\nu_i^{\pm} \geq 0 : i \in \bar{I}\}$ such that the function

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

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

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

Now consider the vector $x_t = w + td$, where $t > 0$. This is a feasible solution to (2.9) due to (c); the $\|\cdot\|_1$-norm of this solution is

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

where the concluding inequality is given by (a) and the fact that $\nu_i^{\pm} \geq 0$. Since $w_i \neq 0$ 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 (2.9) and $\|x_t\|_1 \leq \|w\|_1$ for all small positive t. Since w is the unique optimal solution to (2.9), 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 (A.46) is < 1.

We see that whenever x is a vector with s nonzero entries, equal to ± 1, there exists y such that $(A^T y)_i = x_i$ when $x_i \neq 0$ and $|(A^T y)_i| < 1$ when $x_i = 0$ (indeed, in the role of this vector one can take the y-component of an optimal solution to the problem (A.46) 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 (2.9). There is nothing to prove when $w = 0$. Now let $w \neq 0$, let s' be the number of nonzero entries of w, and I be the set of indices of these entries. By **C** we have $\gamma := \gamma_{s'}(A) \leq \gamma_s(A)$, i.e., $\gamma < 1$. Recalling the definition of $\gamma_s(\cdot)$, there exists $y \in \mathbb{R}^k$ such that $(A^T y)_i = \mathrm{sign}(w_i)$ when $w_i \neq 0$ and $|(A^T y)_i| \leq \gamma$ when $w_i = 0$. The function

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

coincides with the objective of (2.9) on the feasible set of (2.9). Since $|\gamma_i| \leq \gamma < 1$, 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 (2.9). To see that this optimal solution is unique, let x' be another optimal solution to the problem. Then

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

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 *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

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

which is impossible. ∎

(ii) 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

$$\widehat{\gamma}_s\left(A, \frac{1}{1+\gamma}\beta\right) \leq \frac{\gamma}{1+\gamma} < 1/2. \tag{A.47}$$

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

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

where $\bar{I}$ is the complement of I. It is immediately seen that Π_I is a closed and convex set in $\mathbb{R}^n$. Let B be the centered at the origin $\|\cdot\|_\infty$-ball of the radius $1-\widehat{\gamma}$ in $\mathbb{R}^n$: $B = \{u \in \mathbb{R}^n : \|u\|_\infty \le 1-\widehat{\gamma}\}$. We claim that Π_I contains B. 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 $(A^T y)_i = (1-\widehat{\gamma})^{-1} u_i = z_i$ for $i \in I$ and $|(A^T y)_i| \le (1-\widehat{\gamma})^{-1}\widehat{\gamma}$ for $i \notin I$. Thus, the validity of our claim would imply that

$$\widehat{\gamma} := \widehat{\gamma}_s(A,\beta) < 1/2 \Rightarrow \gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\beta\right) \le \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < 1. \tag{A.48}$$

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 $e^T v > \max_{v' \in \bar{Q}} e^T v'$ (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 $\|A^T y - z\|_\infty \le \widehat{\gamma}$; 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

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

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 (A.47), (A.48) which are slightly weakened versions of (2.14.*a*-*b*). It remains to prove that the inequalities $\le$ in the conclusions of (A.47), (A.48) are in fact equalities. This is immediate: assume that under the premise of (2.14.*a*) we have

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

When applying (A.48) with β replaced with $\frac{1}{1+\gamma}\beta$, we get

$$\gamma_s\left(A, \frac{1}{1-\widehat{\gamma}}\left[\frac{1}{1+\gamma}\beta\right]\right) \le \frac{\widehat{\gamma}}{1-\widehat{\gamma}} < \frac{\gamma_+}{1-\gamma_+} = \gamma. \tag{A.49}$$

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 (A.49) implies that $\gamma_s(A,\beta) < \gamma$, which contradicts the definition of γ. Thus, the concluding $\le$ in (A.47) is in fact equality. By completely similar argument, so is the concluding $\le$ in (A.48). ∎

Gaussian matrix

m	lower bounds on $s_*(A)$			upper bound $\overline{s}$	CPU time (s)		
	$s[\mu]$	$s[\alpha_1]$	$s[\alpha_s]$		$s[\alpha_1]$	$s[\alpha_s]$	$\overline{s}$
25	1	1	1	1	11.0	21.6	3.4
51	1	2	2	4	22.3	24.1	8.8
76	1	3	3	7	34.2	34.3	23.1
102	1	3	4	11	50.8	190.7	34.0
128	1	5	5	15	69.3	75.8	31.6
153	1	5	6	19	93.8	557.6	60.7
179	2	7	8	25	115.4	658.3	103.8
204	2	9	11	31	141.2	551.5	97.8
230	2	14	17	41	173.0	561.0	97.8

Random Fourier matrix

m	lower bounds on $s_*(A)$			upper bound $\overline{s}$	CPU time (s)		
	$s[\mu]$	$s[\alpha_1]$	$s[\alpha_s]$		$s[\alpha_1]$	$s[\alpha_s]$	$\overline{s}$
24	1	1	1	2	9.3	6.1	1.3
51	1	2	2	4	129.5	14.5	7.2
76	2	3	3	6	233.1	12.8	16.1
102	2	4	4	7	213.9	11.2	25.6
128	2	4	4	8	270.9	426.5	58.1
152	3	5	5	10	245.9	2350.7	57.8
178	3	6	6	14	319.7	161.2	81.5
204	4	7	7	14	234.0	97.9	75.8
230	4	9	9	19	343.2	76.0	51.9

Random Hadamard matrix

m	lower bounds on $s_*(A)$			upper bound $\overline{s}$	CPU time (s)		
	$s[\mu]$	$s[\alpha_1]$	$s[\alpha_s]$		$s[\alpha_1]$	$s[\alpha_s]$	$\overline{s}$
25	1	1	1	2	10.1	7.4	1.2
51	1	2	2	4	21.6	11.7	3.5
76	2	3	3	4	34.1	14.2	6.7
102	3	4	4	11	50.8	23.8	37.7
128	3	5	5	7	69.6	48.5	24.1
153	3	7	7	11	93.8	31.1	84.7
179	4	9	9	15	112.0	51.0	88.9
204	5	12	12	15	141.6	51.1	78.6
230	6	18	18	28	141.5	55.4	44.1

Table 1: Efficiently computable bounds on $s_*(A)$, $n = 256$.

Lower bound $s[\mu]$: the bound (4.32) based on mutual incoherence; $s[\alpha_1]$-bound: the "improved" bound based on upper bounding of $\alpha_s(A)$ via the matrix Y obtained when computing $\alpha_1(A)$; $s[\alpha_s]$: the bound based on computing $\alpha_s(A)$. **Upper bound** $\overline{s}$: the bound based on successive convex approximation

Gaussian matrix

m	lower bounds on $s_*(A)$		upper bound $\overline{s}$	CPU time (s)	
	$s[\mu]$	$s[\alpha_1]$		$s[\alpha_1]$	$\overline{s}$
102	2	2	8	457.0	400.7
204	2	4	18	1179.0	1722.1
307	2	6	30	2234.6	7585.9
409	3	7	44	3658.6	3421.7
512	3	10	61	5341.7	6304.3
614	3	12	78	7155.7	17616.7
716	3	15	105	9446.1	11670.4
819	4	21	135	12435.1	8373.1
921	4	32	161	13564.2	9838.3

Convolution matrix

m	lower bounds on $s_*(A)$		upper bound $\overline{s}$	CPU time (s)	
	$s[\mu]$	$s[\alpha_1]$		$s[\alpha_1]$	$\overline{s}$
960	0	5	7	4579.1	271.8

Table 2: Efficiently computable bounds on $s_*(A)$, $n = 1024$.

Lower bound $s[\mu]$: the bound (4.32) based on mutual incoherence; $s[\alpha_1]$-bound: the "improved" bound based on upper bounding of $\alpha_s(A)$ via the matrix Y obtained when computing $\alpha_1(A)$. **Upper bound** $\overline{s}$: the bound based on successive convex approximation