Computable Quantification of the Stability of Sparse Signal Reconstruction

Gongguo Tang and Arye Nehorai

Department of Electrical and Systems Engineering
Washington University in St. Louis
St. Louis, MO 63130-1127
Email: nehorai@ese.wustl.edu

***Abstract*—The ℓ_1-constrained minimal singular value (ℓ_1-CMSV) of the sensing matrix is shown to determine, in a concise and tight manner, the recovery performance of ℓ_1-based algorithms such as Basis Pursuit, the Dantzig selector, and the LASSO estimator. Several random measurement ensembles are shown to have ℓ_1-CMSVs bounded away from zero with high probability, as long as the number of measurements is relatively large. Three algorithms based on projected gradient method and interior point algorithm are developed to compute ℓ_1-CMSV. A lower bound of the ℓ_1-CMSV is also available by solving a semi-definite programming problem.**

***Index Terms*—ℓ_1-constrained minimal singular value, Basis Pursuit, Dantzig selector, LASSO estimator, random measurement ensemble, sparse signal reconstruction, semidefinite relaxation**

I. Introduction

Sparse signal reconstruction aims at recovering a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ from observations of the following model:

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ is the measurement or sensing matrix, $\boldsymbol{y}$ is the measurement vector, and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise vector. The sparsity level k of $\boldsymbol{x}$ is defined as the number of non-zero components of $\boldsymbol{x}$. When the sparsity level k is very small, it is possible to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ in a stable manner even if model (1) is underdetermined.

This paper is motivated by the following considerations. When we are given a sensing or measurement system (1), we usually want to know the performance of the system before using it. This involves deriving a performance measure which is computable, given the sensing matrix A as well as the signal and noise structures. Computable performance measures also provide ways to quantify our confidence on the reconstructed signal, especially when we do not have other means to justify the correctness of the recovered signal. Furthermore, in many practical signal processing problems, we usually have the freedom to design the sensing matrix; that is, we can choose the best from a collection of sensing matrices. For example, in radar imaging and sensor array applications, sensing matrix design is connected with waveform design and array configuration design. To optimally design the sensing matrix, we need to

1) analyze how the performance of recovering $\boldsymbol{x}$ from $\boldsymbol{y}$ is affected by A, and define a function $\rho(A)$ to accurately quantify the goodness of A in the context of sparse signal reconstruction;
2) develop algorithms to efficiently compute $\rho(A)$ for arbitrary given A;
3) design mechanisms to select within a matrix class the sensing matrix that is optimal in the sense of best $\rho(A)$.

There is a copious literature on the first aspect but not much on the second and the third. According to the literature (*e.g.*, [1], [2]) the goodness of A is its incoherence property measured by, most notably, the Restricted Isometry Constant (RIC). Unfortunately, for a given arbitrary matrix, the RIC is extremely difficult to compute, if not impossible. So a stochastic framework is usually adopted to argue that the majority of a certain class of randomly generated matrices have good RICs. However, this framework is not satisfactory, considering our three-step scheme for optimal design. As a consequence, the remaining two steps in designing an optimal sensing matrix using RIC are nearly hopeless. This paper investigates the possibility of finding a good measure $\rho(A)$ to fulfill step 1) and possibly step 2). We will not consider step 3) in this paper.

The contribution of the work is threefold. First, we use the ℓ_1-constrained minimal singular value (ℓ_1-CMSV) of the sensing matrix to derive concise bounds on the ℓ_2 norm of estimation error for Basis Pursuit, the Dantzig selector, and the LASSO estimator. Compared with derivations using the RIP condition, the arguments are clearer and more intuitive. Second, we demonstrate that if the number of measurements m is reasonably large, several important random measurement ensembles have ℓ_1-CMSVs bounded away from zero, with high probability. The lower bounds on m are nearly as good as the bounds for the RIP. Last but not least, we develop algorithms to compute the ℓ_1-CMSV for an arbitrary sensing matrix and compare their performance. These algorithms are by no means the most efficient ones. However, once we shift from an optimization problem with a discrete nature (*e.g.*, the RIC) to a continuous one, there are many optimization tools available and more efficient algorithms can be designed.

The paper is organized as follows. In Section II, we present our measurement model, three convex relaxation algorithms, and the RIP condition. Section III is devoted to deriving

This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, and ONR Grant N000140810849.

bounds on the errors of several convex relaxation algorithms. In Section IV, we show that the majority of realizations of several random measurement ensembles have good ℓ_1-CMSVs. In Section V, we design algorithms to compute the ℓ_1-CMSV and the lower bound on it. Section VI summaries our conclusions.

II. Model and Background

Suppose $\boldsymbol{x} \in \mathbb{R}^n$ is a $k-$sparse signal, *i.e.*, $\boldsymbol{x}$ has at most $k \ll n$ non-zero components. We observe $\boldsymbol{y} \in \mathbb{R}^m$ through the following linear model:

$$\boldsymbol{y} \quad = \quad A\boldsymbol{x} + \boldsymbol{w}, \tag{2}$$

where $A \in \mathbb{R}^{m\times n}$ or $\mathbb{C}^{m\times n}$ is the sensing matrix and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise/disturbance vector, either deterministic or random. Two fundamental problems are to reconstruct the signal $\boldsymbol{x}$ from the measurement $\boldsymbol{y}$ by exploiting sparsity, and to quantify the stability of the reconstruction with respect to noise.

We briefly review three reconstruction algorithms based on convex relaxation: Basis Pursuit, the Dantzig selector, and the LASSO estimator. A common theme of these algorithms is enforcing the sparsity of solutions by penalizing large ℓ_1 norms. The Basis Pursuit algorithm [3] tries to minimize the ℓ_1 norm of solutions subject to the measurement constraint:

$$\text{BP}: \quad \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - A\boldsymbol{z}\|_2 \leq \epsilon. \tag{3}$$

The Dantzig selector [4] aims to reconstruct a reasonable signal in most cases when the measurement is contaminated by unbounded noise. Its estimate for $\boldsymbol{x}$ is the solution to the ℓ_1-regularization problem:

$$\text{DS}: \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_1 \quad \text{subject to} \quad \|A^T(\boldsymbol{y} - A\boldsymbol{z})\|_\infty \leq \lambda_n \cdot \sigma, \tag{4}$$

where σ^2 is the noise variance, and λ_n a control parameter. The constraint in (4) requires that all feasible solutions must have uniformly bounded correlations between the induced residual vector and the columns of the sensing matrix.

Following the convention in [5], in this paper we refer to the solution to the following optimization problem as the LASSO estimator:

$$\text{LASSO:} \quad \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{z}\|_1. \tag{5}$$

All three optimizations (3), (4) and (5) can be implemented using convex programming or even linear programming.

The stability analysis of Basis Pursuit, the Dantzig selector, and the LASSO estimator aims to derive error bounds of the solutions of these algorithms. In previous work, these bounds usually involve the incoherence of the sensing matrix A, which is measured by the RIC [2], [6]. For each integer $k \in \{1, \ldots, n\}$, the *restricted isometry constant (RIC)* δ_k of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as the smallest $\delta > 0$ such that

$$1 - \delta \leq \|A\boldsymbol{x}\|_2^2 \leq 1 + \delta \tag{6}$$

holds for arbitrary unit-length $k-$sparse signal $\boldsymbol{x}$.

A matrix A with a small δ_k roughly means that A is nearly an isometry when restricted onto all $k-$sparse vectors. The computation of δ_k involves optimizing over the set of all $k-$sparse signals, which can be divided into $\binom{n}{k}$ subsets with intersections of measure zero by specifying the exact locations of non-zero components. Consequently, the exact computation of the RIC is extremely difficult.

Due to space limitation, we will not explicitly present the error bounds, which can be found in [2], [4], [5]. We comment that these bounds involving the RIC are quite complicated. Their derivations are even more complicated. In addition, the bound for the LASSO estimator requires very technical conditions for its validity.

Although the RIC provides a measure quantifying the goodness of a sensing matrix, as mentioned earlier, its computation poses great challenge. In the literature, the computation issue is circumvented by resorting to a random argument. Here we cite one general result [7]: Let $A \in \mathbb{R}^{m\times n}$ be a random matrix whose entries are *i.i.d.* samples from any subgaussian distribution. Then, for any given $\delta \in (0, 1)$, there exist constants $c_1, c_2 > 0$ depending only on δ such that $\delta_k \leq \delta$, with probability not less than $1 - 2e^{-c_2 m}$, as long as

$$m \geq c_1 k \log(n/k). \tag{7}$$

We remark that subgaussian distributions include the Gaussian distribution and the Bernoulli distribution. For the ℓ_1-CMSV, we will establish a theorem similar to the one above for the Gaussian ensembles. We also present a result for the Bernoulli ensemble and a modified Fourier ensemble with a slightly worse bound on m.

III. Stability of Convex Relaxation based on the ℓ_1-Constrained Minimal Singular Value

In this section, we derive bounds on the reconstruction error for the Basis Pursuit, the Dantzig selector and the LASSO estimator in terms of the ℓ_1-CMSV rather than the RIC of matrix A . We first introduce a quantity that measures the sparsity, (or, more precisely, the density), of a given vector $\boldsymbol{x}$.

Definition 1 *For $p \in (0, 1]$, the ℓ_p-sparsity level of a non-zero vector $\boldsymbol{x} \in \mathbb{R}^n$ is defined as*

$$s_p(\boldsymbol{x}) = \|\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_2, \tag{8}$$

where the ℓ_p semi-norm of $\boldsymbol{x}$ is $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |\boldsymbol{x}_i|^p\right)^{1/p}$.

The scaling and permutation invariant $s_p(\boldsymbol{x})$ is indeed a measure of sparsity. To see this, suppose $\|\boldsymbol{x}\|_0 = k$; then Hölder's inequality implies that $s_p(\boldsymbol{x}) \leq k^{\frac{1}{p}-\frac{1}{2}}$, and we have equality if and only if the absolute values of all non-zero components are equal. Therefore, the more non-zero elements $\boldsymbol{x}$ has and the more evenly the magnitudes of these non-zero elements are distributed, the larger $s_p(\boldsymbol{x})$. In particular, if $\boldsymbol{x}$ has exactly one non-zero element, then $s_p(\boldsymbol{x}) = 1$; if $\boldsymbol{x}$ has n non-zero elements with the same magnitude, then $s_p(\boldsymbol{x}) = n^{1/p-1/2}$. However, if $\boldsymbol{x}$ has n non-zero elements but their magnitudes are spread in a wide range, then its ℓ_p sparsity level might be very small. In this paper, we will focus on the ℓ_1 sparsity level $s_1(\boldsymbol{x})$.

Definition 2 *For any $s \in [1, \sqrt{n}]$ and matrix $A \in \mathbb{R}^{m \times n}$, define the ℓ_1-constrained minimal singular value (abbreviated as ℓ_1-CMSV) and the ℓ_1-constrained maximal singular value of A by*

$$\rho_s^{\min} \stackrel{\text{def}}{=} \min_{\boldsymbol{x} \neq 0,\ s_1(\boldsymbol{x}) \leq s} \|A\boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2, \quad \textit{and} \tag{9}$$

$$\rho_s^{\max} \stackrel{\text{def}}{=} \max_{\boldsymbol{x} \neq 0,\ s_1(\boldsymbol{x}) \leq s} \|A\boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2, \tag{10}$$

respectively. Because we mainly use $\rho_s^{\min}$ in this paper, for notational simplicity, we sometimes use ρ_s to denote $\rho_s^{\min}$ when it causes no confusion.

We first establish a bound on the ℓ_2 norm of Basis Pursuit's error vector using the ℓ_1-CMSV. The procedure of establishing Theorems 1, 2 and 3 has two steps:

1) Show that the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ is ℓ_1-sparse: $s_1(x) \leq 2\sqrt{k}$, which automatically leads to a lower bound $\|A\boldsymbol{h}\|_2 \geq \rho_{2\sqrt{k}} \|\boldsymbol{h}\|_2$;
2) Obtain an upper bound on $\|A\boldsymbol{h}\|_2$.

For Basis Pursuit (3), the second step is trivial as both $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ satisfy the constraint $\|\boldsymbol{y} - A\boldsymbol{z}\| \leq \epsilon$ in (3). The triangle inequality immediately leads to $\|A\boldsymbol{h}\|_2 \leq 2\epsilon$. In order to establish the ℓ_1-sparsity of the error vector in the first step, we use an observation by Candés [2]. The fact that $\|\hat{\boldsymbol{x}}\|_1$ is the minimum among all $\boldsymbol{z}$ satisfying the constraint in (3) implies that $\|\boldsymbol{h}_{S^c}\|_1$ cannot be very large: $\|\boldsymbol{h}_{S^c}\|_1 \leq \|\boldsymbol{h}_S\|_1$, which yields, by the Cauchy-Schwart inequality, $s_1(\boldsymbol{h}) \leq 2\sqrt{k}$.

Consequently, we have the following theorem:

Theorem 1 *If the support of $\boldsymbol{x}$ is of size k and the noise $\boldsymbol{w}$ is bounded; that is, $\|\boldsymbol{w}\|_2 \leq \epsilon$, then the solution $\hat{\boldsymbol{x}}$ to* (3) *obeys*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq 2\epsilon / \rho_{2\sqrt{k}}. \tag{11}$$

Compared with the bound $\frac{4\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \cdot \epsilon$ in [2], the bound $2\epsilon/\rho_{2\sqrt{k}}$ here is more concise, and the argument is simpler. In [2], the major effort is devoted to establishing a lower bound on $\|A\boldsymbol{h}\|_2$, which is trivial according to Definition 2. Of course, if $\rho_{2\sqrt{k}}$ is not bounded away from zero, this concise bound will not offer much. We will show in Section IV that, for several important random measurement ensembles, the induced ℓ_1-CMSVs are bounded away from zero with high probability if m is large enough. Numerical simulations show that randomly generated matrices are more likely to have $\rho_{2\sqrt{k}}$ bounded away from zero than to have $\delta_{2k} < \sqrt{2} - 1$, the prerequisite for the RIC bound $\frac{4\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \cdot \epsilon$ to be applicable. The ℓ_1-CMSV bound is also tighter than the RIC bound when both apply.

We now present similar results for the Dantzig Selector and the LASSO estimator. More details on their proofs and discussions can be found in [8].

Theorem 2 *Suppose $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_m)$ in the sensing model* (2)*, and suppose $\boldsymbol{x} \in \mathbb{R}^n$ is a $k-$sparse signal. Choose $\lambda_n = \sqrt{2 \log n}$ in* (4)*. Then, with high probability, the solution $\hat{\boldsymbol{x}}$ to* (4) *satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq 4\sqrt{2k \log n}\,\sigma / \rho_{2\sqrt{k}}^2. \tag{12}$$

Theorem 3 *Suppose $\boldsymbol{x}$ is a $k-$sparse vector. The LASSO estimator $\hat{\boldsymbol{x}}$ given by the solution to* (5) *in the absence of noise obeys*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \lambda\sqrt{k} / \rho_{2\sqrt{k}}^2. \tag{13}$$

IV. ℓ_1-CONSTRAINED SINGULAR VALUES OF RANDOM MATRICES

This section is devoted to analyzing the property of the ℓ_1-CMSVs for the Gaussian ensemble, the Bernoulli ensemble and a modified Fourier ensemble. We show that, for these sensing matrices, the ℓ_1-CMSVs are bounded away from zero with high probability. We first introduce the notion of Gaussian measurement ensemble:

Definition 3 *A Gaussian sensing matrix $A = [a_{\alpha,\beta}] \in \mathbb{R}^{m \times n}$ is a matrix whose entries follow* i.i.d. *Gaussian distribution with mean* 0 *and variance* $1/m$.

For Gaussian measurement ensemble, we have the following theorem:

Theorem 4 *Suppose $\epsilon \in (0, 1)$. Given $s \in [1, \sqrt{n})$, assume the following condition holds for Gaussian sensing matrix $A \in \mathbb{R}^{m \times n}$:*

$$m \geq cs^2 (\log n)/\epsilon^2 \tag{14}$$

for some absolute constant c. Then, we have

$$\mathbb{P}\{\rho_s < 1 - \epsilon\} \leq \exp(-m\epsilon^2/8). \tag{15}$$

Recall that for any $k-$sparsity signal $\boldsymbol{x} \in \mathbb{R}^n$, its ℓ_1-sparsity level $s_1(\boldsymbol{x}) \leq \sqrt{k}$. Hence, considering the obvious fact that $\delta_k \leq \max\{1 - (\rho_{\sqrt{k}}^{\min})^2, (\rho_{\sqrt{k}}^{\max})^2 - 1\}$, we obtain a corollary of Theorem 5 that the RIC $\delta_k \leq \epsilon$ for Gaussian measurement ensemble with high probability if $m \geq ck \log n/\epsilon^2$, a bound comparable with the one in previous literature [7].

The proof of Theorem 4 employs the comparison theorem for Gaussian random processes, in particular the Gordon's inequality [9], and the concentration of measure phenomena in Gauss space [9], [10]. This argument relies heavily on the symmetry of Gaussian processes and have no obvious generalizations to other distributions. We can not trivially extend the nice $\varepsilon-$net argument in [7] for RIC. One difficulty is the absence of tight covering number estimation for the set of vectors with ℓ_1-sparsity level s. Another difficulty is that the set $\{\boldsymbol{x} \in \mathbb{R}^n : s_1(\boldsymbol{x}) \leq s\}$ can not be partitioned into subsets which are closed under substraction, rendering the approximation using $\epsilon-$cover useless.

The following definitions characterize the Bernoulli measurement ensemble and a modified Fourier measurement ensemble:

Definition 4 *A Bernoulli sensing matrix $A = [a_{\alpha,\beta}] \in \mathbb{R}^{m \times n}$ is produced by generating its entries following independent Bernoulli distributions:*

$$a_{\alpha,\beta} = \begin{cases} +1/\sqrt{m} & \textit{with probability } 1/2, \\ -1/\sqrt{m} & \textit{with probability } 1/2. \end{cases} \tag{16}$$

Definition 5 *Denote* $W_\alpha = [1, e^{-j\frac{2\pi\alpha}{n}}, \ldots, e^{-j\frac{2\pi\alpha(n-1)}{n}}]^T \in \mathbb{C}^n$ *with* $j = \sqrt{-1}$. *A modified Fourier sensing matrix* $A = [\boldsymbol{a}_1^T \cdots \boldsymbol{a}_m^T]^T \in \mathbb{C}^{m\times n}$ *is produced by selecting its* m *rows, with replacement, from* $\{W_0/\sqrt{m}, \ldots, W_{n-1}/\sqrt{m}\}$*:*

$$\mathbb{P}\{\boldsymbol{a}_\alpha = W_\beta/\sqrt{m}\} = 1/n,\ 0 \le \beta < n;\ 1 \le \alpha \le m. \quad (17)$$

Note that in Definition 5, the generated matrix has exactly m rows with possibly repeated rows. In [11] the Fourier sensing matrix selects each of the n Fourier vectors $\{W_0/\sqrt{m}, \ldots, W_{n-1}/\sqrt{m}\}$ with probability $\frac{m}{n}$. Therefore, the size of the resulting matrix is random with approximately m rows in average. It is expected that the behaviors of the two slightly different Fourier ensembles share similar properties in sparse signal reconstruction. For the Bernoulli measurement ensemble and the Fourier measurement ensemble, we have

Theorem 5 *Suppose* $\epsilon \in (0, 1)$. *Given* $s \in [1, \sqrt{n})$*, assume the following condition holds for the Bernoulli measurement ensemble or the modified Fourier measurement ensemble*

$$m \ge Cs^2 \log^3 s \log n (\log\log n)^3/\epsilon^2. \quad (18)$$

Then, we have

$$\mathbb{P}\{\rho_s < 1/2\} \le 6\exp\left(-c/\epsilon^2\right). \quad (19)$$

Here C *and* c *are positive numerical constants.*

The major challenge of establishing Theorem 5 is to derive a good estimate of the (dyadic) entropy number of the operator $A : \ell_1^n \to \ell_2^m$ defined by $\boldsymbol{x} \mapsto A\boldsymbol{x}$. The rest of the proof essentially follows the procedure developed in [12]. For more details, see [13]. Note the lower bound on m and the high probability assertion are worse than the case for Gaussian measurement ensemble. For Bernoulli ensemble, this seems to be a feature of the proof. In any case, by selecting ϵ small enough, we could make $\rho_s > 1/2$ with desired probability.

V. Computation of the ℓ_1-Constrained Singular Values

One major advantage of using the ℓ_1-CMSV as a measure of the "goodness" of a sensing matrix is the relative ease of its computation. We show in this section that $\rho_s^{\min}$ and $\rho_s^{\max}$ are computationally more amenable than δ_k. The discrete nature of δ_k makes the algorithm design for its computation very difficult.

Consider the following optimization problem to obtain the ℓ_1-CMSV, or more precisely, its square:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \boldsymbol{z}^T A^T A \boldsymbol{z} \text{ subject } \|\boldsymbol{z}\|_1 \le s,\ \|\boldsymbol{z}\|_2 = 1. \quad (20)$$

Unfortunately, the above optimization is not convex because of the ℓ_2 constraint $\|\boldsymbol{z}\|_2 = 1$. However, many tools at our disposal can deal with the continuous problem (20), for example, the Lagrange multiplier or the Karush-Kuhn-Tucker condition [14]. In our work [8], we designed three algorithms to directly compute an approximate numerical solution of (20). Because of space limitation, we only present the most effective one based on interior point algorithm. Due to the non-convexity, there is no guarantee that the solutions of these algorithms are the true minima. We will also present a convex program to compute a lower bound on ℓ_1-CMSV.

The interior point (IP) method provides a general approach to efficiently solve general constrained optimization problems. The basic idea is to construct and solve a sequence of penalized optimization problems with equality constraints. The interior point approach assumes objective and constraint functions with continuous second order derivatives, which is not satisfied by the constraint $\|\boldsymbol{z}\|_1 - s \le 0$. To circumvent this difficulty, we define $\boldsymbol{z} = \boldsymbol{z}^+ - \boldsymbol{z}^-$ with $\boldsymbol{z}^+ = \max(\boldsymbol{z}, \boldsymbol{0}) \ge 0$ and $\boldsymbol{z}^- = \max(-\boldsymbol{z}, \boldsymbol{0}) \ge 0$, which leads to the following augmented optimization solved by interior point algorithm:

$$\begin{aligned} \text{IP:} \quad & \min_{\boldsymbol{z}^+, \boldsymbol{z}^- \in \mathbb{R}^n} (\boldsymbol{z}^+ - \boldsymbol{z}^-)^T A^T A (\boldsymbol{z}^+ - \boldsymbol{z}^-) \\ & \text{subject to } \sum_i \boldsymbol{z}_i^+ + \sum_i \boldsymbol{z}_i^- - s \le 0, \\ & \qquad (\boldsymbol{z}^+ - \boldsymbol{z}^-)^T (\boldsymbol{z}^+ - \boldsymbol{z}^-) = 1, \\ & \qquad \boldsymbol{z}^+ \ge 0,\ \boldsymbol{z}^- \ge 0. \end{aligned} \quad (21)$$

We briefly describe a semidefinite relaxation (SDR) scheme to compute a lower bound on ℓ_1-CMSV. A similar approach was employed in [15] to compute an upper bound on sparse variance maximization using the *lifting procedure* for semidefinite programming [16]–[18]. Defining $Z = \boldsymbol{z}^T\boldsymbol{z}$ transforms problem (20) into the following equivalent form:

$$\begin{aligned} \max_{Z\in\mathbb{R}^{n\times n}} \quad & \operatorname{trace}(A^T A Z) \\ \text{subject to} \quad & \mathbf{1}^T |Z| \mathbf{1} \le s^2, \operatorname{trace}(Z) = 1, Z \succeq 0, \\ & \operatorname{rank}(Z) = 1. \end{aligned} \quad (22)$$

The *lifting procedure* relaxes the problem (22) by dropping the $\operatorname{rank}(Z) = 1$ constraint:

$$\begin{aligned} \text{SDR:} \quad & \max_{Z\in\mathbb{R}^{n\times n}} \operatorname{trace}(A^T A Z) \\ & \text{subject to } \mathbf{1}^T |Z| \mathbf{1} \le s^2, \operatorname{trace}(Z) = 1, Z \succeq 0. \end{aligned} \quad (23)$$

Now SDR is a semidefinite programming problem. For a small size problem, a global minimum can be achieved at high precision using SEDUMI, SDPT3 or CVX. However, for relatively large n, the interior point algorithm makes the memory requirement prohibitive (see [15] for more discussion). Although the first order algorithm used in [15] can be adapted to solve a penalized version of SDR, it does not give a direct solution to SDR with fixed s.

We compare the ℓ_1-CMSVs ρ_s and their bounds as a function of s computed by IP and SDR, respectively, for Bernoulli random matrices. First, we consider a small-scale problem with $n = 60$ and $m = 10, 20, 40$. A matrix $B \in \mathbb{R}^{40\times 60}$ with entries $\{+1, -1\}$ following $\frac{1}{2}$ Bernoulli distribution is generated. For $m = 10, 20, 40$, the corresponding Bernoulli matrix A is obtained by taking the first m rows of B. The columns of A are then normalized by multiplying $1/\sqrt{m}$. The normalization implies that $\rho_s \le \rho_1 = 1$. For each m, the sparsity levels s are 20 uniform samples in $[1.5, \sqrt{m/2}]$. The IP uses 30 random initial points and selects the solution

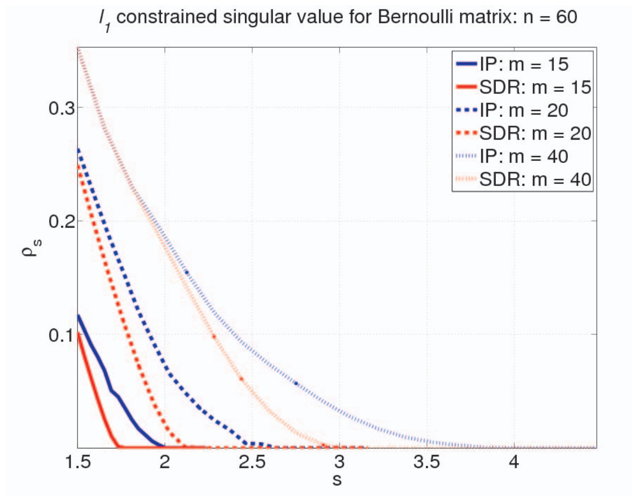

Fig. 1: ℓ_1-CMSV ρ_s and its bound as a function of s for Bernoulli matrix with $n = 60$ and $m = 10, 20, 40$.

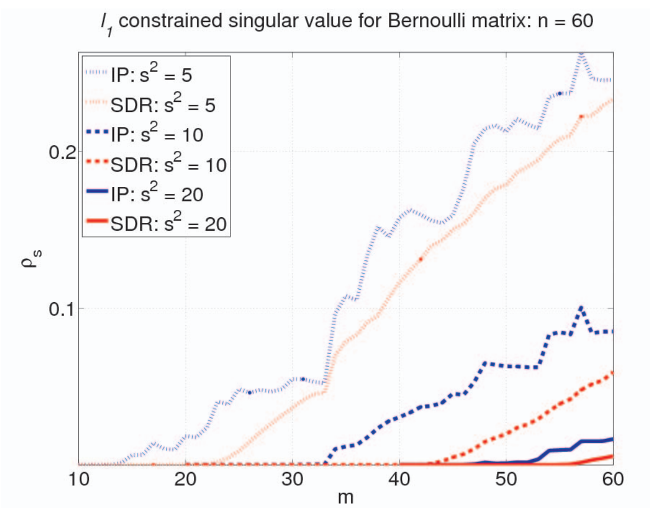

Fig. 2: ℓ_1-CMSV ρ_s and its bound as a function of m for Bernoulli matrix with $n = 60$ and $s = \sqrt{5}, \sqrt{10}, \sqrt{20}$.

with smallest function value. As illustrated in Figure 1, the ℓ_1-CMSVs and their bounds decrease very fast as s increases. For fixed s, increasing m generally (but not necessarily, as shown in Figure 2) increases the ℓ_1-CMSV and their bounds.

In Figure 2, the ℓ_1-CMSV ρ_s is plotted as a function of m with varying parameter values: $s = \sqrt{5}$, $\sqrt{10}$ and $\sqrt{20}$. As in the setup for Figure 1, we first generate a matrix $B \in \mathbb{R}^{60\times 60}$ with entries $\{+1, -1\}$ following $\frac{1}{2}$ Bernoulli distribution. With s fixed, the two algorithms (IP and SDR) are run for $A \in \mathbb{R}^{m\times n}$, with m increasing from $2s^2$ to $n = 60$. For each m, the construction of A follows the procedure described in the previous paragraph. The discrete nature of adding rows to A while increasing m makes the curves in Figure 2 not as smooth as those in Figure 1. The ρ_s increases with m in general, but local decreases do happen. The gap between values computed by IP and SDR is also clearly seen for medium s.

VI. Conclusions

In this paper, a new measure of a sensing matrix's incoherence, the ℓ_1-CMSV, is proposed to quantify the stability of sparse signal reconstruction. It is demonstrated that the reconstruction errors of Basis Pursuit, the Dantzig selector, and the LASSO estimator are concisely bounded using the ℓ_1-CMSV. The ℓ_1-CMSV is shown to be bounded away from zero with high probability for the Gaussian ensemble, the Bernoulli ensemble and the Fourier ensemble, as long as the number of measurements is relatively large. A non-convex program solved by interior point algorithm and one semidefinite program are presented to compute the ℓ_1-CMSV and its lower bound, respectively. The ℓ_1-CMSV provides a computationally amenable measure of incoherence that can be used for optimal design.

References

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, 2008.

[3] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[4] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, pp. 2313–2351, 2007.

[5] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," *Ann. Statist.*, vol. 37, pp. 246–270, 2009.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[7] R. DeVore R. Baraniuk, M. Davenport and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[8] G. Tang and A. Nehorai, "The ℓ_1-constrained minimal singular value: a computable quantification of the stability of sparse signal reconstruction," Submitted to *IEEE Trans. Inf. Theory*.

[9] M. Ledoux and M. Talagrand, *Probability in Banach spaces: Isoperimetry and processes*, vol. 23. of *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 1991.

[10] Michel Ledoux, *The Concentration of Measure Phenomenon*, vol. 89 of *Mathematical Surveys and Monographs*, American Mathematical Society, 2001.

[11] E. J. Candés and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, dec. 2006.

[12] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Proc. 40th Conf. Inf. Sci. Syst. (CISS 2006)*, march 2006, pp. 207 –212.

[13] G. Tang and A. Nehorai, "The ℓ_1 constrained minimal singular value for random matrices," in preparation.

[14] A. Sofer S. G. Nash, *Linear and nonlinear programming*, McGraw-Hill, New York, NY, 1996.

[15] A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, 2007.

[16] L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0-1 optimization," *SIAM Journal on Optimization*, vol. 1, pp. 166–190, 1991.

[17] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM Journal on Optimization*, vol. 5, pp. 13–51, 1995.

[18] C. Lemaréchal and F. Oustry, "Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization," *INRIA, Rapport de recherche*, vol. 3710, 1999.