Performance Analysis of Sparse Recovery Based on Constrained Minimal Singular Values

Gongguo Tang, *Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—The stability of sparse signal reconstruction with respect to measurement noise is investigated in this paper. We design efficient algorithms to verify the sufficient condition for unique ℓ_1 sparse recovery. One of our algorithms produces comparable results with the state-of-the-art technique and performs orders of magnitude faster. We show that the ℓ_1-constrained minimal singular value (ℓ_1-CMSV) of the measurement matrix determines, in a very concise manner, the recovery performance of ℓ_1-based algorithms such as the Basis Pursuit, the Dantzig selector, and the LASSO estimator. Compared to performance analysis involving the Restricted Isometry Constant, the arguments in this paper are much less complicated and provide more intuition on the stability of sparse signal recovery. We show also that, with high probability, the subgaussian ensemble generates measurement matrices with ℓ_1-CMSVs bounded away from zero, as long as the number of measurements is relatively large. To compute the ℓ_1-CMSV and its lower bound, we design two algorithms based on the interior point algorithm and the semidefinite relaxation.**

***Index Terms*—Basis Pursuit, Dantzig selector, interior point algorithm, LASSO estimator, restricted isometry property, semidefinite relaxation, sparse signal reconstruction, verifiable sufficient condition, ℓ_1-constrained minimal singular value.**

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 nonzero components of $\boldsymbol{x}$. The measurement system is underdetermined because the number of measurements m is much smaller than the signal dimension n. However, when the sparsity level k is also small, it is possible to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ in a stable manner. Reconstruction of a sparse signal from linear measurements appears in many signal processing branches, such as compressive sensing [1]–[13], sparse linear regression [4], source localization [5], [6], sparse approximation, and signal denoising [7]. Model (1) is applicable to many practical areas such as DNA microarrays [8], radar imaging [9], cognitive radio [10], and sensor arrays [5], [6], to name a few.

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, or at least to know whether it works in the ideal setting. This involves deriving a verifiable sufficient condition for unique recovery and a computationally amenable performance measure. Furthermore, in signal processing and the applications mentioned in the previous paragraph, 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, respectively. The ℓ_1-CMSV of this paper has already been used to optimally design orthogonal frequency division multiplexing (OFDM) radar signals for detecting a moving target in the presence of multipath reflections [11]. We hope the practitioners in similar fields will also find the results in this paper useful.

The contribution of the work is fourfold. First, we design two algorithms to verify the sufficient condition that a sparse signal can be uniquely recovered using ℓ_1 minimization in the noise-free setting. By solving multiple linear programs efficiently, one of the algorithms produces comparable results with the state-of-the-art verification algorithms and performs orders of magnitude faster. Second, we derive concise bounds on the ℓ_2 norm of estimation error for the Basis Pursuit, the Dantzig selector, and the LASSO estimator in terms of the ℓ_1-CMSV. As the third contribution, we demonstrate that if the number of measurements m is reasonably large, subgaussian random matrices have ℓ_1-CMSVs bounded away from zero, with high probability. This implies that at least for subgaussian random matrices, the ℓ_1-CMSV is as good as the restricted isometry constant. 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 restricted isometry constant) to a continuous one, there are many optimization tools available and more efficient algorithms can be designed.

Many quantities and properties on the sensing matrix A have been proposed to guarantee a stable or unique signal reconstruction, most notably the Restricted Isometry Constant (RIC) [1], [12] and the Null Space Property (NSP) [13]. The RIC provides

Manuscript received February 11, 2011; revised June 28, 2011; accepted July 28, 2011. Date of publication August 15, 2011; date of current version November 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Philippe Ciblat. This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, ONR Grant N000140810849, and NSF Grants CCF-1014908 and CCF-0963742.

The authors are with the Preston M. Green Department of Electrical and Systems Engineering, Washington University, St. Louis, MO 63130 USA (e-mail: gt2@ese.wustl.edu; nehorai@ese.wustl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2164913

a unified framework to deal with sparse signal recovery and has very nice geometrical explanations. However, the RIC is known to be very difficulty to compute. One way to circumvent the computational difficulty is to establish nice probabilistic results of the RIC and the NSP for random sensing matrices [14]–[17]. Another way is to develop computable bounds on quantities involved in the RIC and the NSP using, for example, semi-definite programming relaxation [18], [19], and linear programming relaxation [20]. To the best of the authors' knowledge, the algorithms of [19] and [20] in verifying the sufficient condition of unique ℓ_1 recovery represent the state-of-the-art technique in this direction. In our paper, instead of computing a quantity (e.g., α_k in [20]) for various sparsity levels k and checking if it is less than 1/2, we directly seek the critical sparsity level below which unique recovery is guaranteed. We compare our verification algorithms with those in [19] and [20] using numerical simulations. One of our algorithms performs orders of magnitude faster, consumes much less memory, and produces comparable results.

The paper is organized as follows. In Section II, we present the measurement model, three convex relaxation algorithms, the NSP, and the RIC. Section III is devoted to deriving sufficient conditions for unique ℓ_1 recovery and deriving bounds on the recovery errors of several convex relaxation algorithms. In Section IV, we show that the majority of realizations of the subgaussian measurement ensemble have good ℓ_1-CMSVs. In Section V, we design algorithms to verify unique recovery, and compute the ℓ_1-CMSV and its lower bound. We compare the algorithms' performance in Section VI. Section VII summaries our conclusions.

II. Measurement Model, Reconstruction Algorithms, and Restricted Isometry Property

In this section, we present the measurement model and review three convex relaxation algorithms, the NSP, the RIC, and recovery error bounds based on the RIC.

A. Measurement Model and Recovery Algorithms

The following measurement model is used throughout the paper. Suppose we have a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$, i.e., $\boldsymbol{x}$ has only a few nonzero components. The ℓ_0-sparsity level k of $\boldsymbol{x}$ is defined as the number of nonzero elements of $\boldsymbol{x}$, or the ℓ_0 "norm" of $\boldsymbol{x}$: $k = \|\boldsymbol{x}\|_0$. We call a vector k-sparse if its ℓ_0-sparsity level $\|\boldsymbol{x}\|_0 \leq k$. For ease of presentation, we restrict ourselves to exactly sparse signals in this paper and leave approximately sparse signals to future work.

We observe $\boldsymbol{y} \in \mathbb{R}^m$ through the following linear model:

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

where $A \in \mathbb{R}^{m\times n}$ is the sensing/measurement matrix and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise/disturbance vector.

In this paper, we focus on three most renowned recovery algorithms based on convex relaxation: the Basis Pursuit (BP) [21], the Dantzig Selector (DS) [22], and the LASSO estimator [23]:

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

$$\text{DS}: \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \left\|A^T(\boldsymbol{y} - A\boldsymbol{z})\right\|_\infty \leq \lambda_n\sigma, \quad \text{and} \tag{4}$$

$$\text{LASSO}: \min_{\boldsymbol{z}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 + \lambda_n\sigma\|\boldsymbol{z}\|_1. \tag{5}$$

Here λ_n is a tuning parameter, and ϵ and σ are measures of the noise levels. All three optimization programs can be implemented efficiently using convex programming or even linear programming.

In the noise-free case where $\boldsymbol{w} = 0$, roughly speaking all three algorithms reduce to

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad A\boldsymbol{z} = A\boldsymbol{x}. \tag{6}$$

B. Null Space Property and Restricted Isometry Constant

In the noise-free case, a minimal requirement on the convex relaxation algorithms is the *uniqueness and exactness* of the solution $\hat{\boldsymbol{x}} \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{z}:A\boldsymbol{z}=A\boldsymbol{x}} \|\boldsymbol{x}\|_1$, i.e., $\hat{\boldsymbol{x}} = \boldsymbol{x}$. When the true signal $\boldsymbol{x}$ is k-sparse, the sufficient and necessary condition for unique and exact ℓ_1 recovery is given by the Null Space Property [24]–[26]:

$$\sum_{i\in S} |\boldsymbol{z}_i| < \sum_{i\notin S} |\boldsymbol{z}_i|, \quad \forall \boldsymbol{z} \in \text{null}(A), \quad |S| \leq k \tag{7}$$

where $\text{null}(A) \stackrel{\text{def}}{=} \{\boldsymbol{z} : A\boldsymbol{z} = 0\}$ is the null space of A and $S \subset \{1, \ldots, p\}$ is any index set.

In the noisy case, many quantities are proposed to study the recovery errors of the BP, the DS and the LASSO, for example, the Restricted Isometry Constant (RIC) [3], [12], the Restricted Eigenvalue assumption [27], and the Restricted Correlation assumption [28], among others. However, these quantities are very difficult to compute. For example, the only known technique to exactly compute the RIC is to test all its submatrices of certain sizes.

We will compare our ℓ_1-CMSV based bounds with the RIC based bounds. For this purpose, we follow [3], [12] to define the RIC as follows:

Definition 1: 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

$$1 - \delta \leq \frac{\|A\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \leq 1 + \delta \tag{8}$$

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

The RIC has a very clear geometrical meaning. Roughly speaking, a matrix A with a small δ_k is nearly an isometry when restricted onto all k-sparse vectors.

C. Recovery Error Bounds

Now we cite some of the most renowned performance results on the BP, the DS, and the LASSO, which are expressed in terms of the RIC. Assume $\boldsymbol{x}$ is a k-sparse signal and $\hat{\boldsymbol{x}}$ is its estimate given by any of the three algorithms; then we have the following:

1) BP [12]: Suppose that $\delta_{2k} < \sqrt{2}-1$ and $\|\boldsymbol{w}\|_2 \le \epsilon$. The solution to the BP (3) satisfies

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \frac{4\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \cdot \epsilon. \tag{9}$$

2) DS [22]: If the noise $\boldsymbol{w}$ satisfies $\|A^T\boldsymbol{w}\|_\infty \le \lambda_n\sigma$, and $\delta_{2k}+\delta_{3k}<1$, then, the error signal obeys

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \frac{4\sqrt{k}}{1-\delta_{2k}-\delta_{3k}}\lambda_n\sigma. \tag{10}$$

3) LASSO [29]: If the noise $\boldsymbol{w}$ satisfies $\|A^T\boldsymbol{w}\|_\infty \le \lambda_n\sigma$, and $\delta_{2k} < \frac{1}{(3\sqrt{2}+1)}$, then, the error signal of (5) satisfies

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \frac{16\sqrt{k}}{(1-\delta_{2k})\left(1-\frac{3\sqrt{2}\delta_{2k}}{1-\delta_{2k}}\right)^2}\lambda_n\sigma. \tag{11}$$

We note that in these error bounds, the terms involving the RIC on the right hand sides are quite complicated. We will compare these results with our bounds in Section III, which are much more concise and whose derivations are much less involved.

Although the RIC provides a measure quantifying the goodness of a sensing matrix, as mentioned earlier, its computation poses great challenge. The computation difficulty is compensated by the nice properties of the RIC for a large class of random sensing matrices. We cite one general result below [15]:

- Let $A \in \mathbb{R}^{m\times n}$ be a random matrix whose entries are i.i.d. samples from any distribution that satisfies the concentration inequality for any $\boldsymbol{x}\in\mathbb{R}^n$ and $0<\varepsilon<1$:

$$\mathbb{P}\left(\left|\|A\boldsymbol{x}\|_2^2-\|\boldsymbol{x}\|_2^2\right| \ge \varepsilon\|\boldsymbol{x}\|_2^2\right) \le 2e^{-mc_0(\varepsilon)} \tag{12}$$

where $c_0(\varepsilon)$ is a constant depending only on ε and such that for all $\varepsilon\in(0,1)$, $c_0(\varepsilon)>0$. Then, for any given $\delta\in(0,1)$, there exist constants $c_1,c_2>0$ depending only on δ such that $\delta_k\le\delta$, with probability not less than $1-2e^{-c_2m}$, as long as

$$m \ge c_1 k\log\frac{n}{k}. \tag{13}$$

We remark that distributions satisfying the concentration inequality (12) 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 subgaussian ensemble with almost the same bound on m. The subgaussian ensemble in this paper includes the Gaussian ensemble, the Bernoulli ensemble, as well as the normalized volume measure on various convex symmetric bodies, for example, the unit balls of ℓ_p^n for $2\le p\le\infty$ [30].

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

In this section, we present two approaches to verify the sufficient condition for the uniqueness of ℓ_1-recovery. We also analyze the stability of sparsity recovery with respect to measurement noise by deriving bounds on the reconstruction errors for the BP, the DS and the LASSO. Here by stability we mean small perturbations to the measurement vector $\boldsymbol{y}$ would induce only small changes to the recovered signal. Our bounds are given in terms of the ℓ_1-CMSV rather than the RIC of matrix A.

A. Verifiable Sufficient Conditions for Exact ℓ_1 Recovery

We first introduce a quantity that measures the sparsity, (or, more accurately, the density), of a given vector $\boldsymbol{x}$.

Definition 2: The ℓ_1-sparsity level of a nonzero vector $\boldsymbol{x}\in\mathbb{R}^n$ is defined as

$$s(\boldsymbol{x}) = \frac{\|\boldsymbol{x}\|_1^2}{\|\boldsymbol{x}\|_2^2}. \tag{14}$$

The scaling and permutation invariant $s(\boldsymbol{x})$ is indeed a measure of sparsity. To see this, suppose $\|\boldsymbol{x}\|_0=k$; then the Cauchy-Schwartz inequality implies that

$$s(\boldsymbol{x}) \le k \tag{15}$$

and we have equality if and only if the absolute values of all nonzero components of $\boldsymbol{x}$ are equal. Therefore, the more nonzero elements $\boldsymbol{x}$ has and the more evenly the magnitudes of these nonzero elements are distributed, the larger $s(\boldsymbol{x})$. In particular, if $\boldsymbol{x}$ has exactly one nonzero element, then $s(\boldsymbol{x})=1$; if $\boldsymbol{x}$ has n nonzero elements with the same magnitude, then $s(\boldsymbol{x})=n$.

We use the ℓ_1-sparsity level and its variant as tools to relax the necessary and sufficient condition (7) for exact ℓ_1 recovery in the noise-free setting. More precisely, we have the following sufficient condition for unique and exact ℓ_1 recovery:

Proposition 1: If the sparsity level k is strictly less than either

$$\min_{\boldsymbol{z}:A\boldsymbol{z}=0}\frac{1}{4}\frac{\|\boldsymbol{z}\|_1^2}{\|\boldsymbol{z}\|_2^2}, \text{ or} \tag{16}$$

$$\min_{\boldsymbol{z}:A\boldsymbol{z}=0}\frac{1}{2}\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}, \tag{17}$$

then the optimal solution to (6) is $\boldsymbol{x}$, the true signal.

The proof is given in Appendix A.

Unfortunately, the optimization problem (16) is very difficult to solve exactly. In Section V, we present a semidefinite relaxation algorithm to obtain a lower bound on its optimal value. We will also present a polynomial time algorithm to solve (17). The algorithm solves n linear programs and produces results comparable to the best known results in [20] within a much shorter time.

In the noisy setting, our derivation of the error bounds for the BP, the DS, and the LASSO relies heavily on the fact that the error vectors have small ℓ_1-sparsity levels.

Now we are ready to define the ℓ_1-constrained minimal singular value:

Definition 3: For any real number $s\in[1,n]$ and matrix $A\in\mathbb{R}^{m\times n}$, define the ℓ_1-constrained minimal singular value (abbreviated as ℓ_1-CMSV) of A by

$$\rho_s(A) \stackrel{\text{def}}{=} \min_{\boldsymbol{x}\ne0, s(\boldsymbol{x})\le s}\frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}. \tag{18}$$

Intuitively, the ℓ_1-CMSV $\rho_s(A)$ measures the invertibility of the operator $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ when restricted onto vectors with ℓ_1-sparsity levels not greater than s. For sensing matrices A with columns of unit lengths, we clearly have $\rho_s(A) \leq 1$ because $\|A\mathbf{e}_i\|_2 = 1$ and $s(\mathbf{e}_i) = 1$, where $\mathbf{e}_i$ is the ith canonical basis vector.

In the following theorem, we present our error bounds in terms of the ℓ_1-CMSV, whose proof is given in Appendix B:

Theorem 1: Suppose the support of the true signal $\boldsymbol{x}$ is of size k.

1) If the noise $\boldsymbol{w}$ is bounded; that is, $\|\boldsymbol{w}\|_2 \leq \epsilon$, then the solution $\hat{\boldsymbol{x}}$ to the BS (3) obeys
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2\epsilon}{\rho_{4k}}. \tag{19}$$
2) If the noise $\boldsymbol{w}$ in the DS (4) satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \lambda_n\sigma$, then the solution to (4) obeys
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{k}}{\rho^2_{4k}}\lambda_n\sigma. \tag{20}$$
3) If the noise $\boldsymbol{w}$ in the LASSO (5) satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\lambda_n\sigma$ for some $\kappa \in (0,1)$, then the solution $\hat{\boldsymbol{x}}$ to the LASSO estimator (5) obeys
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{1+\kappa}{1-\kappa} \cdot \frac{2\sqrt{k}}{\rho^2_{\frac{4k}{(1-\kappa)^2}}}\lambda_n\sigma. \tag{21}$$

As shown in Appendix B, the procedure of establishing Theorem 1 has two steps:

1) For all three algorithms, show that the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ is ℓ_1-sparse: $s(\boldsymbol{h}) \leq s$, where $s = 4k$ for the BP and the DS, and $s = \frac{4k}{(1-\kappa)^2}$ for the LASSO. This automatically leads to a lower bound $\|A\boldsymbol{h}\|_2 \geq \rho_s\|\boldsymbol{h}\|_2$;
2) Obtain an upper bound on $\|A\boldsymbol{h}\|_2$ or $\|A\boldsymbol{h}\|_2^2$ and invoke Definition 3 of the ℓ_1-CMSV.

The derivation is simpler than those employed for obtaining the RIC based bounds.

When the noise $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2\mathrm{I}_m)$, as shown by Candés and Tao in [22], with high probability, $\boldsymbol{w}$ satisfies the orthogonality condition
$$|\boldsymbol{w}^T A_j| \leq \lambda_n\sigma \quad \text{for all} \quad 1 \leq j \leq n \tag{22}$$
for $\lambda_n = \sqrt{2\log n}$. More specifically, defining the event
$$E \stackrel{\text{def}}{=} \left\{\|A^T\boldsymbol{w}\|_\infty \leq \lambda_n\sigma\right\} \tag{23}$$
we have
$$\mathbb{P}(E^c) \leq \frac{2n \cdot (2\pi)^{-\frac{1}{2}} e^{\frac{-\lambda_n^2}{2}}}{\lambda_n}. \tag{24}$$
Therefore, with $\lambda_n = \sqrt{2(1+t)\log n}$, we obtain
$$\mathbb{P}(E) \geq 1 - \left(\sqrt{\pi(1+t)\log n} \cdot n^t\right)^{-1}. \tag{25}$$
As a consequence, the conditions on noise in Theorem 1 for the DS and the LASSO hold with high probability.

The ℓ_1-CMSV has the property that $\rho_s(\alpha A) = |\alpha|\rho_s(A)$ for any $\alpha \in \mathbb{R}$. This property together with Theorem 1 imply that increasing the sensing energy without changing the sensing matrix structure reduces the recovery error proportionally. The RIC and the RIC based bounds do not enjoy this property as the RIC is defined on sensing matrices with columns of unit lengths. Furthermore, compared with the RIC bounds (9), (10), and (11), our ℓ_1-CMSV bounds (19), (20), and (21) are more concise. Of course, if the ℓ_1-CMSV $\rho_s(A)$ is not bounded away from zero, these concise bounds would not offer much. We will show in Section IV that, at least for a large class of random matrices, the corresponding ℓ_1-CMSVs are bounded away from zero with high probability if $m \geq c_1 k\log n$. In Section VI, we will use numerical simulations to show that the ℓ_1-CMSV based bounds are tighter and apply to wider ranges of m and k.

IV. ℓ_1-Constrained Minimal Singular Values of Random Matrices

This section is devoted to analyzing the property of the ℓ_1-CMSVs for the subgaussian ensemble. Note that a random vector $\boldsymbol{X} \in \mathbb{R}^n$ is called *isotropic and subgaussian* with constant L if
$$\mathbb{E}\,|\langle \boldsymbol{X}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2, \quad \text{and} \tag{26}$$
$$\mathbb{P}\left(|\langle \boldsymbol{X}, \boldsymbol{u}\rangle| \geq t\right) \leq 2\exp\left(-\frac{t^2}{L\|\boldsymbol{u}\|_2}\right) \tag{27}$$
hold for any $\boldsymbol{u} \in \mathbb{R}^n$. The standard Gaussian vector on $\mathbb{R}^n$ and the sign vector with i.i.d. 1/2 Bernoulli entries are isotropic and subgaussian. Isotropic and subgaussian random vectors also include the vectors with the normalized volume measure on various convex symmetric bodies, for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [30]. In addition, the constants L for these distributions are independent of n. As shown in the following theorem, the ℓ_1-CMSVs associated with subgaussian random matrices are bounded away from zero both in expectation and with high probability:

Theorem 2: Let the rows of the scaled sensing matrix $\sqrt{m}A$ be i.i.d. subgaussian and isotropic random vectors with numerical constant L. Then there exists constants c_1 and c_2 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying
$$m \geq c_1\frac{L^2 s\log n}{\epsilon^2} \tag{28}$$
we have
$$\mathbb{E}\,|1 - \rho_s(A)| \leq \epsilon \tag{29}$$
and
$$\mathbb{P}\left\{1 - \epsilon \leq \rho_s(A) \leq 1 + \epsilon\right\} \geq 1 - \exp\left(-c_2\epsilon^2\frac{m}{L^4}\right). \tag{30}$$

For a detailed proof, see Appendix C. The key to the proof is to reformulate the results of Theorem 2 in term of the empirical process $\sup_{f\in\mathcal{F}}\left|m^{-1}\sum_{k=1}^m f^2(\boldsymbol{a}_i) - \mathbb{E}f^2(\boldsymbol{a})\right|$ and employ some new estimates on the behavior of the empirical process [30, Theorem D]. Relevant concepts and the precise statements of Theorem D of [30] are presented in Appendix C.

Theorem 2 states that at least for subgaussian ensembles (including the Gaussian ensemble and the Bernoulli ensemble), the ℓ_1-CMSV bounds are as tight as the RIC bounds. In Fig. 1, we plot the histogram of $\rho_s(A)$ approximated using the algorithm

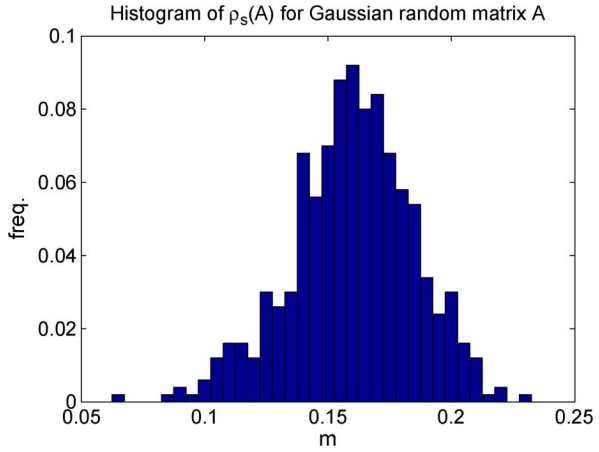

Fig. 1. Histogram of the ℓ_1-CMSV for Gaussian matrices of size 40×60.

(41) for Gaussian random matrices $A \in \mathbb{R}^{40\times 60}$. We set the parameter $s = 4$ and create the histogram from $N = 500$ Gaussian matrices. We observe that the ℓ_1-CMSVs in this particular experiment are indeed bounded away from zero both in expectation and with high probability. We comment that, due to the unknown of the explicit values of the numerical constants in Theorem 2, it is difficult to directly verify Theorem 2 using Monte Carlo simulations.

Following the proof procedure in [16] and using several novel entropy number estimates, we can also establish a result similar to (but slightly worse than) that of Theorem 2 for bounded orthonormal systems [31], which include the partial Fourier matrix and Hadamard matrix as important special cases. Due to space limitation, we choose to not include this result in the current paper. Interested readers can refer to the technical note [32] for more details.

V. Computation of the ℓ_1-CMSVs

In this section, we first describe two algorithms to compute a lower bound on the maximal k such that the sufficient condition (7) is satisfied. This gives a way to verify that the ℓ_1 recovery is exact in the noiseless setting. The second part of this section is devoted to the computation of the ℓ_1-CMSV and its lower bound.

A. Verifying Sufficient Conditions for Exact ℓ_1 Recovery

Using the ℓ_1-sparsity level to verify the sufficient condition (7) (refer to Section III) is formulated as the following optimization problem:

$$\min_{\boldsymbol{z}:A\boldsymbol{z}=0} \frac{1}{4}\frac{\|\boldsymbol{z}\|_1^2}{\|\boldsymbol{z}\|_2^2} \tag{31}$$

or equivalently,

$$\max_{\boldsymbol{z}} 4\|\boldsymbol{z}\|_2^2 \quad \text{s.t.} \quad A\boldsymbol{z} = 0, \quad \|\boldsymbol{z}\|_1 \le 1. \tag{32}$$

Unfortunately, this later optimization problem, which maximizes the ℓ_2 norm over a polyhedron, is NP-hard [33]. By defining $Z = \boldsymbol{z}\boldsymbol{z}^T$ and dropping the rank constraint, we instead use the following semidefinite relaxation to produce an upper bound:

$$(\mathrm{L}_2): \quad \max_{Z:Z\succeq 0} 4\mathrm{trace}(Z) \quad \text{s.t. } \mathrm{trace}(AZA^T) = 0, \quad \|Z\|_1 \le 1, \tag{33}$$

where $\|Z\|_1$ is the sum of absolute values of all elements in Z.

Another relaxation based on the ℓ_∞ norm is to solve the following optimization problem (refer to (17)):

$$(\mathrm{L}_\infty): \quad \max_{\boldsymbol{z}} 2\|\boldsymbol{z}\|_\infty \text{ s.t. } A\boldsymbol{z} = 0, \quad \|\boldsymbol{z}\|_1 \le 1. \tag{34}$$

We write (34) more explicitly as follows

$$\max_{\left\{\begin{array}{c}\boldsymbol{z}: A\boldsymbol{z}=0\\ \|\boldsymbol{z}\|_1\le 1\end{array}\right\}} \max_{1\le i\le n} 2|\boldsymbol{z}_i| = \max_{1\le i\le n} \max_{\left\{\begin{array}{c}\boldsymbol{z}: A\boldsymbol{z}=0\\ \|\boldsymbol{z}\|_1\le 1\end{array}\right\}} 2|\boldsymbol{z}_i| = \max_{1\le i\le n} \max_{\left\{\begin{array}{c}\boldsymbol{z}: A\boldsymbol{z}=0\\ \|\boldsymbol{z}\|_1\le 1\end{array}\right\}} 2\boldsymbol{z}_i, \tag{35}$$

where for the first equality we have exchanged the two maximizations, and for the second equality we have used the fact that the domain for the second maximization is symmetric to the origin. As a consequence, the optimal value of (34) is obtained by solving the following n linear programs:

$$\max_{\boldsymbol{z}} 2\boldsymbol{z}_i \quad \text{s.t.} \quad A\boldsymbol{z} = 0, \quad \|\boldsymbol{z}\|_1 \le 1, \quad i = 1, \ldots, n. \tag{36}$$

The dual problem of the ith subprogram in (36) is

$$2\min_{\boldsymbol{\lambda}_i} \|\mathbf{e}_i - A^T\boldsymbol{\lambda}_i\|_\infty \tag{37}$$

which is actually the subproblem used in [20] to compute $\alpha_1(A,\beta)$ when $\beta = \infty$. In this paper, the linear program subproblems in (36) are implemented using the primal-dual algorithm detailed in [34, Chapter 11]. In [20], Juditsky and Nemirovski also used some techniques to improve the results obtained through (37) that we do not pursue in this paper. Our primal-dual implementation produces results comparable to those in [20] but is significantly faster than the implementations of [20] using the commercial LP solver `mosek` [35] and the non-Euclidean mirror-prox algorithm.

B. Computing the ℓ_1-CMSV

An advantage of using the ℓ_1-CMSV as a measure of the "goodness" of a sensing matrix is the *relative* ease of its computation. The computation of the ℓ_1-CMSV is equivalent to

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \le \sqrt{s}, \quad \|\boldsymbol{x}\|_2 = 1. \tag{38}$$

Unfortunately, the above optimization is not convex because of the ℓ_2 constraint $\|\boldsymbol{x}\|_2 = 1$. However, many tools at our disposal can deal with the continuous problem (38), for example, the Lagrange multiplier or the Karush-Kuhn-Tucker condition [36]. We present an interior point algorithm to directly compute an approximate numerical solution of (38). Since the optimization

problem (38) is not convex, there is no guarantee that the solutions of the algorithm are the true minima. Thus, we also present a convex program to compute a lower bound on the ℓ_1-CMSV.

The interior point (IP) method provides a general approach to efficiently solve the following general constrained optimization problem:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} F(\boldsymbol{z}) \quad \text{s.t.} \quad f(\boldsymbol{z}) \le 0, \quad g(\boldsymbol{z}) = 0. \tag{39}$$

The basic idea is to construct and solve a sequence of penalized optimization problems with equality constraints:

$$\begin{aligned} &\min_{\boldsymbol{z},\sigma} F(\boldsymbol{z}) - \mu \sum_i \log(\sigma_i) \\ &\text{s.t.} \quad f(\boldsymbol{z}) + \sigma = 0, \quad g(\boldsymbol{z}) = 0. \end{aligned} \tag{40}$$

By using either a Newton step, which tries to solve the Karush-Kuhn-Tucker equations [36], or a conjugate gradient step using trust regions to solve the penalized problem (40) in each iteration, the interior point approach efficiently generates a sequence of solutions that converge to the solution of (39). Refer to [37]–[39] for more information on this approach.

However, the interior point approach assumes that the objective and constraint functions have continuous second order derivatives, which is not satisfied by the constraint $\|\boldsymbol{z}\|_1 - \sqrt{s} \le 0$. We address the nondifferentiability of $f(\boldsymbol{z}) = \|\boldsymbol{z}\|_1 - \sqrt{s}$ by defining $\boldsymbol{z} = \boldsymbol{z}^+ - \boldsymbol{z}^-$ 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:

$$\begin{aligned} \text{IP}: \quad &\min_{\boldsymbol{z}^+,\boldsymbol{z}^-\in\mathbb{R}^n} (\boldsymbol{z}^+ - \boldsymbol{z}^-)^T A^T A (\boldsymbol{z}^+ - \boldsymbol{z}^-) \\ \text{subject to} \quad &\sum_i \boldsymbol{z}_i^+ + \sum_i \boldsymbol{z}_i^- - \sqrt{s} \le 0, \\ &(\boldsymbol{z}^+ - \boldsymbol{z}^-)^T (\boldsymbol{z}^+ - \boldsymbol{z}^-) = 1, \\ &\boldsymbol{z}^+ \ge 0, \quad \boldsymbol{z}^- \ge 0. \end{aligned} \tag{41}$$

This algorithm is employed in [11] to design the transmitting waveform of an OFDM radar for optimal detection and estimation performance.

We briefly describe a semidefinite relaxation (SDR) approach to compute a lower bound on the ℓ_1-CMSV. A similar method was employed in [18] to compute an upper bound on sparse variance maximization using the *lifting procedure* for semidefinite programming [40], [41]. Defining $Z = \boldsymbol{z}\boldsymbol{z}^T$ and dropping the rank constraint transform problem (38) into the following form:

$$\begin{aligned} \text{SDR}: \quad &\min_{Z\succeq 0} \quad \operatorname{trace}(A^T A Z) \\ &\text{s.t.} \quad \|Z\|_1 \le s, \quad \operatorname{trace}(Z) = 1. \end{aligned} \tag{42}$$

The semidefinite relaxation (42) yields a bona fide *lower bound* on the optimal ℓ_1-CMSV because the minimization is now taken over a larger feasible set. Therefore, if we replace the ℓ_1-CMSV in Theorem 1 with the semidefinite approximation, the resulting bounds are bona fide upper bounds on sparse recovery errors.

Now the SDR is a semidefinite programming problem. For a small size problem, a global minimum can be achieved at high precision using SEDUMI [42], SDPT3 [43], or CVX [44]. However, for relatively large n, the interior point algorithm makes the memory requirement prohibitive (see [18] for more discussion). In this paper, we do not consider more efficient implementations of the SDR.

TABLE I
COMPARISON OF DIFFERENT VERIFICATION ALGORITHMS FOR A BERNOULLI MATRIX WITH $n = 40$

m	lower bounds on k^*				CPU time (s)			
	L_2	L_∞	d'AE	JN	L_2	L_∞	d'AE	JN
20	1	1	2	1	14.29	0.41	1040.40	4.59
24	1	2	2	2	16.11	0.38	694.20	0.69
28	2	3	3	3	15.12	0.37	710.90	16.20
32	2	3	4	3	15.43	0.37	894.08	2.45

TABLE II
COMPARISON OF L_∞ AND JN FOR A HADAMARD MATRIX WITH $n = 256$

m	lower bounds on k^*		CPU time (s)	
	L_∞	JN	L_∞	JN
25	1	1	3	35
51	2	2	6	70
76	3	3	7	102
102	4	4	9	303
128	5	5	9	544
153	7	7	13	310
179	9	9	15	528
204	12	12	18	1333
230	19	18	18	435

VI. NUMERICAL SIMULATIONS

We use numerical simulations to assess the effectiveness and efficiency of the algorithms presented in Section V. Except for the JN algorithm in Table IV, all other experiments were performed on a platform with a Pentium D CPU@3.40 GHz, 2 GB RAM, and a Windows XP operating system.

A. Verification of Sufficient Conditions

We first examine the L_2 (33) and L_∞ (34) algorithms for verifying the sufficient condition (7). These two algorithms are compared with the two algorithms d'AE and JN proposed in [19] and [20], respectively. We name the two algorithms d'AE and JN using the abbreviations of the authors' names. Define k^* as the maximal k such that (7) is satisfied. In Table I, we show the lower bounds on k^* for a small size Bernoulli matrix with $n = 40$ computed by L_2, L_∞, d'AE and JN. The algorithms of d'AE and JN are provided by the authors for free download online.

In the next set of experiments, we compare lower bounds on k^* computed by L_∞ and JN, respectively, for $n = 256$. In this case, both the semidefinite relaxation in this paper and that in [19] are too time and memory consuming to compute. The lower bounds and execution times are shown in Tables II and III for a Hadamard matrix and a Gaussian matrix, respectively.

Table IV shows the results of L_∞ and JN for a Hadamard matrix with $n = 1024$. Note the lower bounds computed by JN and the CPU times of JN in Table IV are extracted from [20]. We were not able to carry out the computation of JN within reasonable time in our platform.

From Table I, we see that for $n = 40$, d'AE performs the best, and L_∞ and JN give exactly the same results. However, d'AE is very slow in general. For example, even with a first

TABLE III
COMPARISON OF L_∞ AND JN FOR A GAUSSIAN MATRIX WITH $n = 256$

m	lower bounds on k^*		CPU time (s)	
	L_∞	JN	L_∞	JN
25	1	1	6	91
51	2	2	8	191
76	3	3	10	856
102	4	4	13	5630
128	4	5	16	5711
153	6	6	20	1381
179	7	7	24	3356
204	10	10	25	10039
230	13	14	28	8332

TABLE IV
COMPARISON OF L_∞ AND JN FOR GAUSSIAN AND HADAMARD MATRICES WITH $n = 1024$. IN THE COLUMN HEAD, "G" REPRESENTS GAUSSIAN MATRIX AND "H" REPRESENTS HADAMARD MATRIX

m	lower bounds on k^*			CPU time (s)		
	L_∞(H)	L_∞(G)	JN(G)	L_∞(H)	L_∞(G)	JN(G)
102	3	2	2	182	136	457
204	4	4	4	501	281	1179
307	6	6	6	872	510	2235
409	8	7	7	1413	793	3659
512	11	10	10	1914	990	5348
614	14	12	12	1362	1309	7156
716	18	15	15	1687	1679	9446
819	24	20	21	1972	2033	12435
921	37	29	32	2307	2312	13564

order implementation, the d'AE takes more than 37 hours for matrices of size 350×500, while the L_∞ takes less than 3 minutes to finish the computation. From Tables II, III and IV, we see that L_∞ and JN produces comparable results. However, our L_∞ algorithm performs much faster than the JN algorithm. Because the two algorithms solve n linear program subproblems that are dual to each other, they should yield exactly the same results. However, we observe that sometimes the upper bound and lower bound on *the lower bound of* k^* computed by JN does not coincide. The difference in speed might come from the implementation. Our implementation of the linear sub-program employs the primal-dual approach detailed in [34] while [20] uses the commercial LP solver `mosek` [35].

B. Computation of the ℓ_1-CMSV

We next report the test results of the IP and SDR algorithms for computing the ℓ_1-CMSV and its lower bound, respectively. The interior point algorithms IP is implemented using the MATLAB function *fmincon*. The SDR is solved using CVX [44] with the default SDPT3 solver.

We first test the IP and the SDR on a Gaussian matrix $A \in \mathbb{R}^{20\times 60}$ for $s = 5$. Due to the existence of local minima, we need to run the IP several times and select the minimal function value among all the trials as the ℓ_1-CMSV. Fifty random initial points on the unit sphere in $\mathbb{R}^{60}$ are generated for the IP. The SDR only runs once. The results are shown in Table V. In this example, the SDR over-relaxes the problem and produces a zero ℓ_1-CMSV.

We compare the ℓ_1-CMSVs ρ_s and its bound as a function of s approximated by the IP and the SDR, respectively, for Bernoulli random matrices. 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 to have unit norms. The normalization implies that $\rho_s \le \rho_1 = 1$. The IP uses 30 random initial points. As illustrated in Fig. 2, 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 Fig. 3) increases the ℓ_1-CMSV and their bounds.

In Fig. 3, the ℓ_1-CMSV ρ_s is plotted as a function of m with varying parameter values: $s = 5$, 10 and 20. With s fixed, the two algorithms (IP and SDR) are run for $A \in \mathbb{R}^{m\times n}$, with m increasing from $2s$ 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 Fig. 3 not as smooth as those in Fig. 2. The ρ_s increases with m in general, but local decreases do happen. The

TABLE V
COMPARISON OF THE IP AND THE SDR FOR A GAUSSIAN MATRIX

	$\min F(z^*)$	mean$F(z^*)$	std$F(z^*)$	mean time (s)
IP	0.0666	0.7133	0.3661	144.5150
SDR	0.0000	0.0000	N/A	53.1583

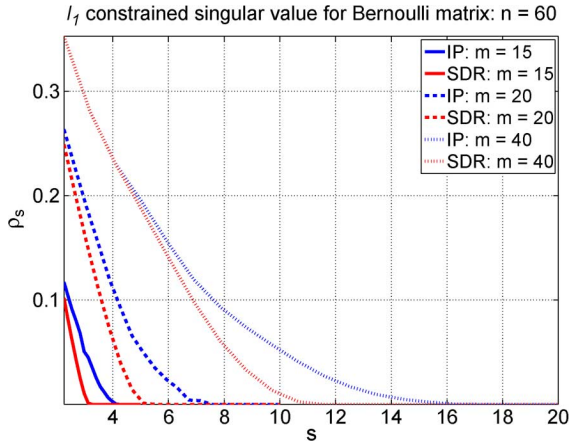

Fig. 2. ℓ_1-CMSV ρ_s and its bound as a function of s for Bernoulli matrices with $n = 60$ and $m = 10, 20, 40$.

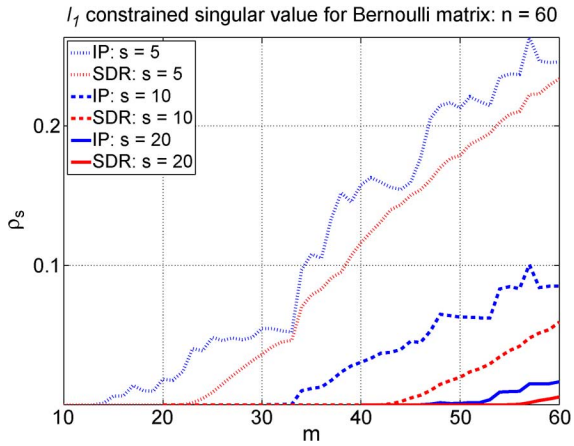

Fig. 3. ℓ_1-CMSV ρ_s and its bound as a function of m for Bernoulli matrices with $n = 60$ and $s = 5, 10, 20$.

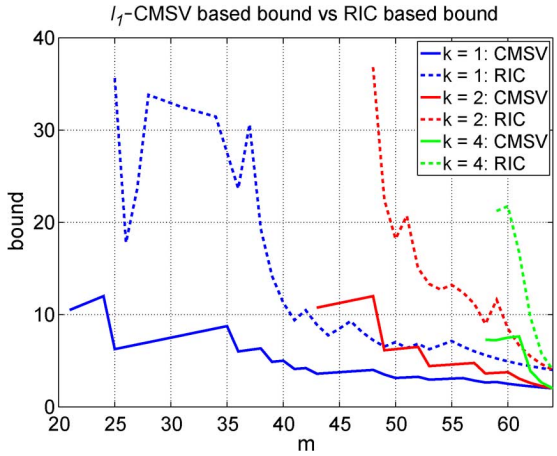

Fig. 4. The ℓ_1-CMSV based bound vs the RIC based bound for Hadamard submatrices with $n = 64$ and $k = 1, 2, 4$.

gap between values computed by the IP and the SDR is also clearly seen for medium s.

In the last set of experiments, for different configurations of m and k, we compare the ℓ_1-CMSV based bound (19) and the RIC based bound (9) on the BP. Without loss of generality, we set the parameter $\epsilon = 1$. The RIC in bound (9) is approximated using Monte Carlo simulations. More explicitly, for $\delta_{2k}(A)$, we randomly take 1000 submatrices of $A \in \mathbb{R}^{m\times n}$ of size $m \times 2k$, compute the maximal and minimal singular values σ_1 and σ_{2k}, and approximate $\delta_{2k}(A)$ using the maximum of $\max(\sigma_1^2 - 1, 1 - \sigma_{2k}^2)$ among all sampled sub-matrices. Obviously, the approximated RIC is always smaller than or equal to the exact RIC.

We approximate the ℓ_1-CMSV and the RIC for column-normalized submatrices of a row-randomly-permuted Hadamard matrix with $n = 64$, $k = 1, 2, 4$ and $m = 10k : n$. As shown in Fig. 4, for all tested cases the ℓ_1-CMSV based bounds are smaller than the RIC based bounds. In addition, for certain configurations of m and k, the ℓ_1-CMSV based bounds applies (i.e., $\rho_{4k}(A) > 0$ and the corresponding finite bound (19) is shown in Fig. 4) even when the RIC based bounds do not apply (i.e., $\delta_{2k}(A) \geq \sqrt{2} - 1$ and the bound (9) does not apply and is not shown in Fig. 4). When m approaches n, we observe that the ℓ_1-CMSV based bounds approach 2 while the RIC based bounds approach 4. This coincides with the limits of bounds (19) and (9) when $\rho_s(A) \to 1$ and $\delta_{2k}(A) \to 0$, respectively. We conclude that for the test cases the ℓ_1-CMSV based bounds are tighter and apply to wider ranges of m and k.

VII. 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 the Basis Pursuit, the Dantzig selector, and the LASSO estimator are concisely bounded using the ℓ_1-CMSV. We also show that the ℓ_1-CMSV is bounded away from zero with high probability for the subgaussian ensemble, as long as the number of measurements is relatively large. One interior point program and one semidefinite program are presented to compute the ℓ_1-CMSV and its lower bound, respectively. Numerical simulations assess the algorithms' performance. The ℓ_1-CMSV provides a computationally amenable goodness measure of the sensing matrix that can be used for optimal design.

As a by product, two algorithms are designed to verify the sufficient conditions guaranteeing the uniqueness of ℓ_1-based recovery. The ℓ_∞ relaxation based algorithm is shown to produce comparable results with the state-of-the-art algorithms, and performs much faster.

Appendix A
Proof of Proposition 1

Proof: Since the negation of (7):

$$\exists \boldsymbol{z} \in \text{null}(A) \text{ and } S \text{ with size at most } k \text{ such that } \sum_{i\in S} |\boldsymbol{z}_i| \geq \sum_{i\notin S} |\boldsymbol{z}_i|$$

implies

$$\begin{aligned} \|\boldsymbol{z}\|_1 \leq & 2\sum_{i\in S} |\boldsymbol{z}_i| \\ \leq & 2\sqrt{k}\sqrt{\sum_{i\in S} |\boldsymbol{z}_i|^2} \\ \leq & 2\sqrt{k}\|\boldsymbol{z}\|_2 \end{aligned}$$

or equivalently

$$k \geq \min_{\boldsymbol{z}:A\boldsymbol{z}=0} \frac{1}{4}\frac{\|\boldsymbol{z}\|_1^2}{\|\boldsymbol{z}\|_2^2}. \tag{43}$$

As a consequence of contraposition, k less than the optimal value of (16) implies the NSP (7), and hence the uniqueness and exactness of the solution to (6).

Another relaxation approach is to replace the ℓ_2 norm in the definition of the ℓ_1-sparsity level with the ℓ_∞ norm. Note that the negation of (7) also implies that

$$\|\boldsymbol{z}\|_1 \leq 2\sum_{i\in S} |\boldsymbol{z}_i| \tag{44}$$

$$\leq 2k\|\boldsymbol{z}\|_\infty. \tag{45}$$

Therefore, the following optimization problem:

$$\min_{\boldsymbol{z}:A\boldsymbol{z}=0} \frac{1}{2}\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty} \tag{46}$$

also provides a way to verify the uniqueness and exactness of the solution to (6). ■

Appendix B
Proof of Theorem 1

In this appendix, we derive the error bounds presented in Theorem 1.

Proof of Theorem 1: We strictly follow the two-step procedure expounded in Section III.

1) In order to establish the ℓ_1-sparsity of the error vector in the first step, we suppose $S = \text{supp}(\boldsymbol{x})$ and $|S| = \|\boldsymbol{x}\|_0 = k$. Define the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$. For any vector $\boldsymbol{z} \in \mathbb{R}^n$ and any index set $S \subseteq \{1, \ldots, n\}$, we use $\boldsymbol{z}_S \in \mathbb{R}^{|S|}$ to represent the vector whose elements are those of $\boldsymbol{z}$ indicated by S.

We first deal with the BP and the DS. As observed by Candés in [12], the fact that $||\hat{\boldsymbol{x}}||_1 = ||\boldsymbol{x}+\boldsymbol{h}||_1$ is the minimum among all $\boldsymbol{z}$ satisfying the constraints in (3) and (4), *together with* the fact that the true signal $\boldsymbol{x}$ satisfies the constraints as required by the conditions imposed on the noise in Theorem 1, imply that $||\boldsymbol{h}_{S^c}||_1$ cannot be very large. To see this, we observe that

$$\begin{aligned}||\boldsymbol{x}||_1 &\geq ||\boldsymbol{x}+\boldsymbol{h}||_1\\ &= \sum_{i\in S}|\boldsymbol{x}_i+\boldsymbol{h}_i| + \sum_{i\in S^c}|\boldsymbol{x}_i+\boldsymbol{h}_i|\\ &\geq ||\boldsymbol{x}_S||_1 - ||\boldsymbol{h}_S||_1 + ||\boldsymbol{h}_{S^c}||_1\\ &= ||\boldsymbol{x}||_1 - ||\boldsymbol{h}_S||_1 + ||\boldsymbol{h}_{S^c}||_1. \end{aligned} \tag{47}$$

Therefore, we obtain $||\boldsymbol{h}_{S^c}||_1 \leq ||\boldsymbol{h}_S||_1$, which leads to

$$\begin{aligned}||\boldsymbol{h}||_1 &= ||\boldsymbol{h}_S||_1 + ||\boldsymbol{h}_{S^c}||_1\\ &\leq 2||\boldsymbol{h}_S||_1\\ &\leq 2\sqrt{k}||\boldsymbol{h}_S||_2\\ &\leq 2\sqrt{k}||\boldsymbol{h}||_2 \end{aligned} \tag{48}$$

where for the next to the last inequality we used the Cauchy-Schwart inequality. Inequality (48) is equivalent to

$$s(\boldsymbol{h}) \leq 4k. \tag{49}$$

We now continue to establish the ℓ_1 sparsity of the error vector for the LASSO (5). We borrow ideas from [45] (see also [27]). Since the noise $\boldsymbol{w}$ satisfies $||A^T\boldsymbol{w}||_\infty \leq \kappa\lambda_n\sigma$ for some small $\kappa > 0$ and $\hat{\boldsymbol{x}}$ is a solution to (5), we have

$$\frac{1}{2}||A\hat{\boldsymbol{x}}-\boldsymbol{y}||_2^2 + \lambda_n\sigma||\hat{\boldsymbol{x}}||_1 \leq \frac{1}{2}||A\boldsymbol{x}-\boldsymbol{y}||_2^2 + \lambda_n\sigma||\boldsymbol{x}||_1. \tag{50}$$

Consequently, substituting $\boldsymbol{y} = A\boldsymbol{x}+\boldsymbol{w}$ yields

$$\begin{aligned}\lambda_n\sigma||\hat{\boldsymbol{x}}||_1 &\leq \frac{1}{2}||A\boldsymbol{x}-\boldsymbol{y}||_2^2 - \frac{1}{2}||A\hat{\boldsymbol{x}}-\boldsymbol{y}||_2^2 + \lambda_n\sigma||\boldsymbol{x}||_1\\ &= \frac{1}{2}||\boldsymbol{w}||_2^2 - \frac{1}{2}||A(\hat{\boldsymbol{x}}-\boldsymbol{x})-\boldsymbol{w}||_2^2 + \lambda_n\sigma||\boldsymbol{x}||_1\\ &= \frac{1}{2}||\boldsymbol{w}||_2^2 - \frac{1}{2}||A(\hat{\boldsymbol{x}}-\boldsymbol{x})||_2^2\\ &\quad + \langle A(\hat{\boldsymbol{x}}-\boldsymbol{x}),\boldsymbol{w}\rangle - \frac{1}{2}||\boldsymbol{w}||_2^2 + \lambda_n\sigma||\boldsymbol{x}||_1\\ &\leq \langle A(\hat{\boldsymbol{x}}-\boldsymbol{x}),\boldsymbol{w}\rangle + \lambda_n\sigma||\boldsymbol{x}||_1\\ &= \langle \hat{\boldsymbol{x}}-\boldsymbol{x}, A^T\boldsymbol{w}\rangle + \lambda_n\sigma||\boldsymbol{x}||_1. \end{aligned} \tag{51}$$

Using the Cauchy-Swcharz type inequality, we get

$$\begin{aligned}\lambda_n\sigma||\hat{\boldsymbol{x}}||_1 &\leq ||\hat{\boldsymbol{x}}-\boldsymbol{x}||_1||A^T\boldsymbol{w}||_\infty + \lambda_n\sigma||\boldsymbol{x}||_1\\ &= \kappa\lambda_n\sigma||\boldsymbol{h}||_1 + \lambda_n\sigma||\boldsymbol{x}||_1, \end{aligned} \tag{52}$$

which leads to

$$||\hat{\boldsymbol{x}}||_1 \leq \kappa||\boldsymbol{h}||_1 + ||\boldsymbol{x}||_1. \tag{53}$$

Therefore, similar to the argument in (47) we have

$$\begin{aligned}||\boldsymbol{x}||_1 &\geq ||\hat{\boldsymbol{x}}||_1 - \kappa||\boldsymbol{h}||_1\\ &= ||\boldsymbol{x}+\boldsymbol{h}_{S^c}+\boldsymbol{h}_S||_1 - \kappa(||\boldsymbol{h}_{S^c}+\boldsymbol{h}_S||_1)\\ &\geq ||\boldsymbol{x}+\boldsymbol{h}_{S^c}||_1 - ||\boldsymbol{h}_S||_1 - \kappa(||\boldsymbol{h}_{S^c}||_1 + ||\boldsymbol{h}_S||_1)\\ &= ||\boldsymbol{x}||_1 + (1-\kappa)||\boldsymbol{h}_{S^c}||_1 - (1+\kappa)||\boldsymbol{h}_S||_1, \end{aligned} \tag{54}$$

where $S = \mathrm{supp}(\boldsymbol{x})$. Consequently, we have

$$||\boldsymbol{h}_{S^c}||_1 \leq \frac{1+\kappa}{1-\kappa}||\boldsymbol{h}_S||_1. \tag{55}$$

Therefore, an argument similar to the one leading to (48) yields

$$||\boldsymbol{h}||_1 \leq \frac{2}{1-\kappa}\sqrt{k}||\boldsymbol{h}||_2 \tag{56}$$

or equivalently,

$$s(\boldsymbol{h}) \leq \frac{4k}{(1-\kappa)^2}. \tag{57}$$

2) We now turn to obtain an upper bound on $||A\boldsymbol{h}||_2$. For the BP (3), this is trivial because both $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ satisfy constraint $||\boldsymbol{y}-A\boldsymbol{z}|| \leq \epsilon$ in (3). The triangle inequality yields

$$\begin{aligned}||A\boldsymbol{h}||_2 &= ||A(\hat{\boldsymbol{x}}-\boldsymbol{x})||_2\\ &\leq ||A\hat{\boldsymbol{x}}-\boldsymbol{y}||_2 + ||\boldsymbol{y}-A\boldsymbol{x}||_2\\ &\leq 2\epsilon. \end{aligned} \tag{58}$$

It then follows from Definition 3 that

$$\rho_{4k}||\boldsymbol{h}||_2 \leq ||A\boldsymbol{h}||_2 \leq 2\epsilon. \tag{59}$$

Hence, we get

$$||\hat{\boldsymbol{x}}-\boldsymbol{x}||_2 \leq \frac{2\epsilon}{\rho_{4k}}. \tag{60}$$

For the DS (4), as shown in [22], the condition on noise $||A^T\boldsymbol{w}||_\infty \leq \lambda_n\sigma$ and the constraint in the Dantzig selector (4) yield

$$||A^TA\boldsymbol{h}||_\infty \leq 2\lambda_n\sigma \tag{61}$$

because

$$\begin{aligned}A_j^T(\boldsymbol{w}-\hat{\boldsymbol{r}}) &= A_j^T\left[(\boldsymbol{y}-A\boldsymbol{x})-(\boldsymbol{y}-A\hat{\boldsymbol{x}})\right]\\ &= A_j^T(A\hat{\boldsymbol{x}}-A\boldsymbol{x}) = A_j^TA\boldsymbol{h}, \end{aligned} \tag{62}$$

where $\hat{\boldsymbol{r}} = \boldsymbol{y}-A\hat{\boldsymbol{x}}$ is the residual corresponding to the Dantzig selector solution $\hat{\boldsymbol{x}}$. Therefore, we obtain an upper bound on $||A\boldsymbol{h}||_2^2$ as follows:

$$\begin{aligned}\boldsymbol{h}^TA^TA\boldsymbol{h} &= \left|\sum_{i=1}^n \boldsymbol{h}_i(A^TA\boldsymbol{h})_i\right|\\ &\leq \sum_{i=1}^n |\boldsymbol{h}_i|\cdot\left|(A^TA\boldsymbol{h})_i\right|\\ &\leq 2\lambda_n\sigma||\boldsymbol{h}||_1. \end{aligned} \tag{63}$$

Equation (63), the definition of ρ_{4k}, and (48) together yield

$$\begin{aligned}\rho_{4k}^2\|\boldsymbol{h}\|_2^2 &\le \boldsymbol{h}^T A^T A\boldsymbol{h}\\ &\le 2\lambda_n\sigma\|\boldsymbol{h}\|_1\\ &\le 4\lambda_n\sqrt{k}\sigma\|\boldsymbol{h}\|_2. \end{aligned}\tag{64}$$

We conclude that

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \frac{4\sqrt{k}}{\rho_{4k}^2}\lambda_n\sigma. \tag{65}$$

Now we establish an upper bound on $\|A\boldsymbol{h}\|_2^2$ for the LASSO (5) using a procedure similar to the one used for the DS given above. We need to establish a bound on

$$\begin{aligned}\|A^TA\boldsymbol{h}\|_\infty &\le \left\|A^T(\boldsymbol{y}-A\boldsymbol{x})\right\|_\infty + \left\|A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}})\right\|_\infty\\ &\le \|A^T\boldsymbol{w}\|_\infty + \left\|A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}})\right\|_\infty\\ &= \kappa\lambda_n\sigma + \left\|A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}})\right\|_\infty. \end{aligned}\tag{66}$$

We again follow the procedure in [45] (see also [27]) to estimate $\|A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}})\|_\infty$. Since $\hat{\boldsymbol{x}}$ is the solution to (5), the optimality condition yields that

$$A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}}) \in \lambda_n\sigma\partial\|\hat{\boldsymbol{x}}\|_1 \tag{67}$$

where $\partial\|\hat{\boldsymbol{x}}\|_1 = [-1,1]^n$ is the subgradient of $\|\cdot\|_1$ evaluated at $\hat{\boldsymbol{x}}$.

As a consequence, we obtain

$$\left\|A^T(\boldsymbol{y}-A\hat{\boldsymbol{x}})\right\|_\infty \le \lambda_n\sigma. \tag{68}$$

Following the same lines in (63), we get

$$\|A\boldsymbol{h}\|_2^2 \le (\kappa+1)\lambda_n\sigma\|\boldsymbol{h}\|_1. \tag{69}$$

Then, (56), (66), and (68)

$$\begin{aligned}\rho^2_{\frac{4k}{(1-\kappa)^2}}\|\boldsymbol{h}\|_2^2 &\le \|A\boldsymbol{h}\|_2^2\\ &\le (\kappa+1)\lambda_n\sigma\frac{\sqrt{4k}}{1-\kappa}\|\boldsymbol{h}\|_2. \end{aligned}\tag{70}$$

As a consequence, we get

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \frac{1+\kappa}{1-\kappa}\cdot\frac{2\sqrt{k}}{\rho^2_{\frac{4k}{(1-\kappa)^2}}}\lambda_n\sigma. \tag{71}$$

■

Appendix C
Proof of Theorem 2

The key to the proof of Theorem 2 is some new estimates on the behavior the empirical process $\sup_{f\in\mathcal{F}}|m^{-1}\sum_{k=1}^m f^2(\boldsymbol{a}_i) - \mathbb{E}f^2(\boldsymbol{a})|$ established in [30, Theorem D]. For a better understanding of the precise statement of Theorem D in [30], we first present some relevant concepts and notations. For any scalar random variable X, the Orlicz ψ_2 norm is defined as

$$\|X\|_{\psi_2} = \inf\left\{t>0 : \mathbb{E}\exp\left(\frac{|X|^2}{t^2}\right)\le 2\right\}. \tag{72}$$

Markov's inequality immediately gives that X with finite $\|X\|_{\psi_2}$ has subgaussian tail

$$\mathbb{P}\left(|X|\ge t\right) \le 2\exp\left(\frac{-t^2}{\|X\|_{\psi_2}}\right). \tag{73}$$

The converse is also true, i.e., if X has subgaussian tail $\exp\left(\frac{-t^2}{K^2}\right)$, then $\|X\|_{\psi_2}\le cK$. In viewing this relationship between the subgaussian tail and the Orlicz ψ_2 norm, we conclude that a random vector $\boldsymbol{X}\in\mathbb{R}^n$ is *isotropic and subgaussian* with constant L if $\mathbb{E}|\langle\boldsymbol{X},\boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\|\langle\boldsymbol{X},\boldsymbol{u}\rangle\|_{\psi_2}\le L\|\boldsymbol{u}\|_2$ hold for any $\boldsymbol{u}\in\mathbb{R}^n$, a definition equivalent to that given by (26) and (27) with properly adjusted constant L.

Another important concept in studying empirical process is the ℓ_*-functional defined here.

Definition 4: Let $\mathcal{H}\subset\mathbb{R}^n$ and $\boldsymbol{g}\sim\mathcal{N}(0,\mathrm{I}_n)$. Denote by $\ell_*(\mathcal{H}) = \mathbb{E}\sup_{\boldsymbol{u}\in\mathcal{H}}\langle\boldsymbol{g},\boldsymbol{u}\rangle$.

With these preparations, we combine [30, Theorem D] and the comments afterwards to present the following theorem:

Theorem 3 [30, Theorem D]: Let $\{\boldsymbol{a},\boldsymbol{a}_i, i=1,\dots,m\}\subset\mathbb{R}^n$ be i.i.d. isotropic and subgaussian random vectors, $\mathcal{H}$ is a subset of the unit sphere of $\mathbb{R}^n$, and $\mathcal{F} = \{f_{\boldsymbol{u}}(\cdot) = \langle\boldsymbol{u},\cdot\rangle : \boldsymbol{u}\in\mathcal{H}\}$. Suppose $\mathrm{diam}(\mathcal{F},\|\cdot\|_{\psi_2}) \overset{\text{def}}{=} \max_{f,g\in\mathcal{F}}\|f-g\|_{\psi_2} = \alpha$. Then there exist absolute constants c_1,c_2,c_3 such that for any $\epsilon>0$ and $m\ge 1$ satisfying

$$m \ge c_1\frac{\alpha^2\ell_*^2(\mathcal{H})}{\epsilon^2} \tag{74}$$

with probability at least $1-\exp\left(\frac{-c_2\epsilon^2 m}{\alpha^4}\right)$,

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^m f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right| \le \epsilon. \tag{75}$$

Furthermore, if $\mathcal{F}$ is symmetric, we have

$$\begin{aligned}&\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^m f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right|\\ &\qquad\le c_3\max\left\{\alpha\frac{\ell_*(\mathcal{H})}{\sqrt{m}}, \frac{\ell_*^2(\mathcal{H})}{m}\right\}. \end{aligned}\tag{76}$$

Now we are ready to prove Theorem 2.

Proof: We connect the ℓ_1-CMSV for the sensing matrix A with an empirical process. Suppose the rows of $\sqrt{m}A$ are i.i.d. isotropic and subgaussian random vectors with constant L, and are denoted by $\{\boldsymbol{a}_i^T, i=1,\dots,m\}$. Denote $\mathcal{H}_s^n = \{\boldsymbol{u}\in\mathbb{R}^n : \|\boldsymbol{u}\|_2^2 = 1, \|\boldsymbol{u}\|_1^2\le s\}$, a subset of the unit sphere of $\mathbb{R}^n$. We observe that for any $\epsilon\in(0,1)$

$$\rho_s(A)^2 = \min_{\boldsymbol{u}:\boldsymbol{u}\in\mathcal{H}_s^n}\boldsymbol{u}^TA^TA\boldsymbol{u} < (1-\epsilon)^2 < (1-\epsilon) \tag{77}$$

is a consequence of

$$\begin{aligned}&\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\left|\frac{1}{m}\boldsymbol{u}^T(\sqrt{m}A)^T(\sqrt{m}A)\boldsymbol{u}-1\right|\\ &\qquad= \sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\left|\frac{1}{m}\sum_{i=1}^m\langle\boldsymbol{a}_i,\boldsymbol{u}\rangle^2 - 1\right| \le \epsilon. \end{aligned}\tag{78}$$

Define a class of functions parameterized by $\boldsymbol{u}$ as $\mathcal{F}_s = \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot \rangle : \boldsymbol{u} \in \mathcal{H}_s^n\}$. Since $\mathbb{E}f^2(\boldsymbol{a}) = \mathbb{E}\langle \boldsymbol{u}, \boldsymbol{a}\rangle^2 = \|\boldsymbol{u}\|_2^2 = 1$ due to isotropy, the proof of Theorem 2 boils down to estimating

$$\mathbb{E}\sup_{f\in\mathcal{F}_s}\left|\frac{1}{m}\sum_{k=1}^{m} f^2(\boldsymbol{a}_i) - \mathbb{E}f^2(\boldsymbol{a})\right| \tag{79}$$

and

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}_s}\left|\frac{1}{m}\sum_{k=1}^{m} f^2(\boldsymbol{a}_i) - \mathbb{E}f^2\right|\right\} \tag{80}$$

using Theorem 3. The symmetry of $\mathcal{H}_s^n$ (and hence of $\mathcal{F}_s$) yields

$$\begin{aligned}\alpha &= \operatorname{diam}(\mathcal{F}_s, |\cdot|_{\psi_2}) \\ &= 2\sup_{\boldsymbol{u}\in\mathcal{H}_s^n} \|\langle \boldsymbol{u}, \boldsymbol{a}\rangle\|_{\psi_2} \le 2L.\end{aligned} \tag{81}$$

Now the key is to compute $\ell_*(\mathcal{H}_s^n)$ (actually an upper bound suffices). Clearly, we have

$$\begin{aligned}\ell_*(\mathcal{H}_s^n) &= \mathbb{E}\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\langle \boldsymbol{g}, \boldsymbol{u}\rangle \\ &\le \mathbb{E}\|\boldsymbol{u}\|_1\|\boldsymbol{g}\|_\infty \\ &\le \sqrt{s\log n}.\end{aligned} \tag{82}$$

The conclusions of Theorem 2 then follow from (82) and Theorem 3 with suitable choice of c_1. ■

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

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

[4] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 451–460, Feb. 2007.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[6] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Process.*, vol. 86, pp. 624–638, Mar. 2006.

[7] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[8] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, Jun. 2008.

[9] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[10] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2007)*, Honolulu, HI, Apr. 2007, pp. IV-1357–IV-1360.

[11] S. Sen, G. Tang, and A. Nehorai, "Multiobjective optimization of OFDM radar waveform for target detection," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 639–652, Feb. 2011.

[12] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," in *Compte Rendus de l'Acad. Sci., Paris, Serie I*, 2008, vol. 346, pp. 589–592.

[13] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, Jul. 2009.

[14] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," in *Proc. Nat. Acad. Sci. USA*, 2005, vol. 102, no. 27, pp. 9452–9457.

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

[16] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure and Appl. Math.*, vol. 61, pp. 1025–1045, 2008.

[17] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing-probabilistic analysis of a null-space characterization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2008)*, Las Vegas, NV, Mar. 2008, pp. 3377–3380.

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

[19] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Math. Program.*, vol. 127, no. 1, pp. 123–144, 2011.

[20] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization," *Math. Program.*, vol. 127, no. 1, pp. 57–88, 2011.

[21] 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.

[22] 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.

[23] R. Tibshirani, "Regression shrinkage and selection via LASSO," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 267–288.

[24] Y. Zhang, A Simple Proof for Recoverability of ℓ_1-Minimization: Go Over or Under? Rice CAAM Dep. , 2005, Tech. Rep..

[25] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[26] D. Donoho, High-Dimensional Centrally-Symmetric Polytopes With Neighborliness Proportional to Dimension Dep. Statist., Stanford Univ., Stanford, CA, 2004.

[27] P. Bickel, Y. Ritov, and A. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Statist.*, vol. 37, no. 4, pp. 1705–1732, 2009.

[28] P. J. Bickel, "Discussion of the Dantzig selector: Statistical estimation when p is much larger than n, by E. J. Candès and T. Tao," *Ann. Stat.*, vol. 35, pp. 2352–2357, 2007.

[29] V. Kekatos and G. B. Giannakis, Sparse Volterra and Polynomial Regression Models: Recoverability and Estimation Arxiv preprint arXiv:1103.0769, 2011.

[30] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and sub-Gaussian operators in asymptotic geometric analysis," *Geometr. Funct. Analysis.*, pp. 1248–1282, Nov. 2007.

[31] H. Rauhut, "Compressive sensing and structured random matrices," *Theoret. Found. Numer. Methods for Sparse Recov.*, vol. 9, pp. 1–92, 2010.

[32] G. Tang, "The ℓ_1-CMSV for bounded orthonormal systems," Dept. Elect. Syst. Eng., Washington Univ., St. Louis, MO, Tech. Note, 2011.

[33] H. Bodlaender, P. Gritzmann, V. Klee, and J. Leeuwen, "Computational complexity of norm-maximization," *Combinatorica*, vol. 10, pp. 203–225, 1990.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[35] E. Andersen and K. Andersen, The MOSEK Optimization Tools Manual, Version 5.0 [Online]. Available: http://www.mosek.com

[36] A. Sofer and S. G. Nash, *Linear and Nonlinear Programming*. New York: McGraw-Hill, 1996.

[37] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Math. Program.*, vol. 89, no. 1, pp. 149–185, 2000.

[38] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optimiz.*, vol. 9, no. 4, pp. 877–900, 1999.

[39] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Math. Program.*, vol. 107, no. 3, pp. 391–408, 2006.

[40] L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0–1 optimization," *SIAM J. Optimiz.*, vol. 1, pp. 166–190, 1991.
[41] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM J. Optimiz.*, vol. 5, pp. 13–51, 1995.
[42] J. Sturm, "Using SEDUMI 1.0 ×, a MATLAB toolbox for optimization over symmetric cones," *Optimiz. Methods and Software*, vol. 11, pp. 625–653, 1999.
[43] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Math. Program. Ser. B*, vol. 95, pp. 189–217, 2003.
[44] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming (Web Page and Software). Jun. 2009.
[45] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.

Gongguo Tang (S'09–M'11) received the B.Sc. degree in mathematics from the Shandong University, China, in 2003, and the M.Sc. degree in systems science from Chinese Academy of Sciences, in 2006, and the Ph.D. degree in electrical and systems engineering from Washington University, St. Louis, MO, in 2011.

He is currently a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Wisconsin-Madison. His research interests are in the area of sparse signal processing, matrix completion, mathematical programming, statistical signal processing, detection and estimation, and their applications.

Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees from the Technion, Israel, and the Ph.D. degree from Stanford University, CA.

He is the Eugene and Martha Lohman Professor and Chair of The Preston M. Green Department of Electrical and Systems Engineering, Washington University, St. Louis (WUSTL). He serves as the Director of the Center for Sensor Signal and Information Processing at WUSTL. Prior to that, he was a faculty member at Yale University, New Haven, CT, and the University of Illinois at Chicago.

Dr. Nehorai served as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2000–2002. He was Vice President (Publications) of the IEEE Signal Processing Society (SPS), Chair of the Publications Board, and member of the Executive Committee of this Society during 2003–2005. He was the Founding Editor of the special columns on Leadership Reflections in the *IEEE Signal Processing Magazine* from 2003 to 2006. He received the 2006 IEEE SPS Technical Achievement Award and the 2010 IEEE SPS Meritorious Service Award. He was elected Distinguished Lecturer of the IEEE SPS for the term 2004 to 2005. He was a corecipient of the IEEE SPS 1989 Senior Award for Best Paper, a coauthor of the 2003 Young Author Best Paper Award, and a corecipient of the 2004 Magazine Paper Award. In 2001, he was named University Scholar of the University of Illinois. He has been a Fellow of the Royal Statistical Society since 1996.