Computable Performance Bounds on Sparse Recovery

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

***Abstract*—In this paper, we develop verifiable sufficient conditions and computable performance bounds of ℓ_1-minimization based sparse recovery algorithms in both the noise-free and noisy cases. We define a family of quality measures for arbitrary sensing matrices as a set of optimization problems, and design polynomial-time algorithms with theoretical global convergence guarantees to compute these quality measures. The proposed algorithms solve a series of second-order cone programs, or linear programs. We derive performance bounds on the recovery errors in terms of these quality measures. We also analytically demonstrate that the developed quality measures are non-degenerate for a large class of random sensing matrices, as long as the number of measurements is relatively large. Numerical experiments show that, compared with the restricted isometry based performance bounds, our error bounds apply to a wider range of problems and are tighter, when the sparsity levels of the signals are relatively low.**

***Index Terms*—Compressive sensing, computable performance bounds, linear programming, second-order cone programming, sparse recovery.**

I. Introduction

SPARSE signal recovery (particularly compressive sensing) has revolutionized the way we think of signal sampling [1]. It goes far beyond sampling and has also been applied to areas as diverse as medical imaging, remote sensing, radar, sensor arrays, image processing, computer vision, and so on. Mathematically, sparse signal recovery aims to reconstruct a sparse signal, namely a signal with only a few non-zero components, from usually noisy linear measurements:

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

where $\boldsymbol{x} \in \mathbb{R}^n$ is the sparse signal, $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement vector, $A \in \mathbb{R}^{m\times n}$ is the sensing/measurement matrix, and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise. A theoretically justified way to exploit the sparseness in recovering $\boldsymbol{x}$ is to minimize its ℓ_1 norm under certain constraints [2].

In this paper, we investigate the problem of using the ℓ_∞ norm as a performance criterion for sparse signal recovery via ℓ_1 minimization. Although the ℓ_2 norm has been used as the performance criterion by the majority of published research in sparse signal recovery, the adoption of the ℓ_∞ norm is well justified. Other popular performance criteria, such as the ℓ_1 and ℓ_2 norms of the error vectors, can all be expressed in terms of the ℓ_∞ norm in a tight and non-trivial manner. More importantly, the ℓ_∞ norm of the error vector has a direct connection with the support recovery problem. To see this, assuming we know *a priori* the minimal non-zero absolute value of the components of the sparse signal, then controlling the ℓ_∞ norm within half of that value would guarantee exact recovery of the support. Support recovery is arguably one of the most important and challenging problems in sparse signal recovery. In practical applications, the support is usually physically more significant than the component values. For example, in radar imaging using sparse signal recovery, the sparsity constraints are usually imposed on the discretized time-frequency domain. The distance and velocity of a target have a direct correspondence to the support of the sparse signal. In this example, *a priori* knowledge of the minimal non-zero magnitude of the target is a realistic assumption. The magnitudes determined by coefficients of reflection are of less physical significance [3]–[5]. Refer to [6] for more discussions on sparse support recovery.

Another, perhaps more important, reason to use the ℓ_∞ norm as a performance criterion is the computability of the resulting performance bounds. A general strategy to study the performance of sparse signal recovery is to define a measure of the quality of the sensing matrix, and then derive performance bounds in terms of the quality measure. The most well-known quality measure is the restricted isometry constant (RIC) [7]. Upper bounds on the ℓ_2 and ℓ_1 norms of the error vectors for various recovery algorithms have been expressed in terms of the RIC. Unfortunately, it is NP-hard to compute the RIC. The only known sensing matrices with nice RICs are certain types of random matrices [8]. Many other types of conditions guaranteeing exact or approximate sparse recovery were proposed in the literature [9]–[11], however, except for the mutual coherence, which typically leads to very loose performance bounds, none of these conditions can be verified efficiently.

By using the ℓ_∞ norm as a performance criterion, we develop a framework in which a family of quality measures for

Manuscript received December 13, 2013; revised June 15, 2014; accepted September 06, 2014. Date of publication October 29, 2014; date of current version December 03, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hongwei Liu. This work was supported by AFOSR Grant No. FA9550-11-1-0210, NSF Grants CCF-1014908 and CCF-0963742, and ONR Grant No. N000141310050. A preliminary version of this work appeared in "Verifiable and Computable ℓ_∞ Performance Evaluation of ℓ_1 Sparse Signal Recovery," in the *Proceedings of the Forty-Fifth Annual Conference on Information Sciences and Systems* 2011.

G. Tang is with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401 USA (e-mail: gtang@mines.edu).

A. Nehorai is with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: 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.2014.2365766

the sensing matrices can be computed exactly in polynomial-time. We achieve the computability by sacrificing the tightness of the error bounds. In applications where strict performance guarantee is preferable, such as radar and sensor arrays, these loose bounds provide more meaningful guidance in practice than bounds that are not computable at all. The computability further justifies the connection of the ℓ_∞ norm with the support recovery problem, since for the connection described in the previous paragraph to be practically useful, we must be able to compute the error bounds on the ℓ_∞ norm. The verifiability and computability therefore open doors for wide applications. In many practical applications of sparse signal recovery, e.g., radar imaging [12], sensor arrays [13], DNA microarrays [14], and MRI [15], it is beneficial to know the performance of the sensing system before implementation and taking measurements.

We now preview our contributions. First of all, we use the ℓ_∞ norm as a performance criterion for sparse signal recovery and establish its connections with other performance criteria. We define a family of quality measures of the sensing matrix, and use them to derive performance bounds on the ℓ_∞ norm of the recovery error vector. Performance bounds using other norms are expressed using the ℓ_∞ norm. Although there are sporadic previous work that derives error bound using the ℓ_∞ norm [16], our systematic treatment identifies its fundamental role in establishing computable performance bounds. Secondly, we develop polynomial-time algorithms to efficiently compute the quality measures for given sensing matrices by solving a series of second-order cone programs or linear programs, depending on the specific quality measure being computed. As a by-product, we obtain a fast algorithm to verify the sufficient condition guaranteeing exact sparse recovery via ℓ_1 minimization. The verifiable sufficient condition is generally much weaker than those based on mutual coherence. Finally, we show that the quality measures are non-degenerate for subgaussian and isotropic random sensing matrices as long as the number of measurements is relatively large, a result parallel to but worse than that of the RIC for random matrices.

Several attempts have been made to verify the Null Space Property (NSP) [17] and compute the RIC [2], [7]. Since both problems are NP-hard [18], [19], researchers use relaxation techniques to obtain an approximate solution. Examples include semi-definite programming relaxation [20], [21] and linear programming relaxation [8]. To the best of the authors' knowledge, the algorithms of [8] and [21] represent state-of-the-art techniques in verifying the sufficient condition of unique ℓ_1 recovery. In this paper, we directly address the computability of the performance bounds. More explicitly, we define the quality measures as the optimal values of some optimization problems and design efficient algorithms with theoretical convergence guarantees to solve the optimization problems. An algorithm to verify a sufficient condition for exact ℓ_1 recovery is obtained as a by-product. Our implementation of the algorithm performs orders of magnitude faster than the state-of-the-art techniques in [8] and [21], consumes much less memory, and produces comparable results.

The paper is organized as follows. In Section II, we introduce notations, and present the measurement model, three convex relaxation algorithms, and the sufficient and necessary condition for exact ℓ_1 recovery. In Section III, we derive performance bounds on the ℓ_∞ norms of the recovery errors for several convex relaxation algorithms. In Section IV, we design algorithms to verify a sufficient condition for exact ℓ_1 recovery in the noise-free case, and to compute the quality measures of arbitrarily given sensing matrices. Section V is devoted to the probabilistic analysis of our ℓ_∞ performance measures. We evaluate the algorithms' performance in Section VI. Section VII summarizes our conclusions.

II. Notations, Measurement Model, and Recovery Algorithms

In this section, we introduce notations and the measurement model, and review sparse recovery algorithms based on ℓ_1 minimization.

For any vector $\boldsymbol{x} \in \mathbb{R}^n$, the norm $\|\boldsymbol{x}\|_{k,1}$ is the summation of the absolute values of the k (absolutely) largest components of $\boldsymbol{x}$. In particular, the ℓ_∞ norm $\|\boldsymbol{x}\|_\infty = \|\boldsymbol{x}\|_{1,1}$ and the ℓ_1 norm $\|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_{n,1}$. The classical inner product in $\mathbb{R}^n$ is denoted by $\langle \cdot, \cdot \rangle$, and the ℓ_2 (or Euclidean) norm is $\|\boldsymbol{x}\|_2 = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. We use $\|\cdot\|_\diamond$ to denote a general norm.

The support of $\boldsymbol{x}$, $\mathrm{supp}(\boldsymbol{x})$, is the index set of the non-zero components of $\boldsymbol{x}$. The size of the support, usually denoted by the ℓ_0 "norm" $\|\boldsymbol{x}\|_0$, is the sparsity level of $\boldsymbol{x}$. Signals of sparsity level at most k are called $k-$sparse signals. If $S \subset \{1, \cdots, n\}$ is an index set, then $|S|$ is the cardinality of S, and $\boldsymbol{x}_S \in \mathbb{R}^{|S|}$ is the vector formed by the components of $\boldsymbol{x}$ with indices in S.

We use $\mathbf{e}_i$, $\mathbf{0}$, $\boldsymbol{O}$, and $\mathbf{1}$ to denote respectively the ith canonical basis vector, the zero column vector, the zero matrix, and the column vector with all ones.

Suppose $\boldsymbol{x}$ is a k-sparse signal. In this paper, we observe $\boldsymbol{x}$ through the following linear model:

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

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

Many algorithms have been proposed to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ by exploiting the sparseness of $\boldsymbol{x}$. We focus on three algorithms based on ℓ_1 minimization: the Basis Pursuit Denoising [22], the Dantzig selector [23], and the LASSO estimator [24].

$$\text{Basis Pursuit Denoising}: \min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } \|\boldsymbol{y} - A\boldsymbol{z}\|_\diamond \le \varepsilon \tag{3}$$

$$\text{Dantzig}: \min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } \left\|A^T(\boldsymbol{y} - A\boldsymbol{z})\right\|_\infty \le \mu \tag{4}$$

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

Here μ is a regularization parameter, and ε is a measure of the noise level. All three optimization problems have efficient implementations.

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

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

which is the ℓ_1 relaxation of the NP hard ℓ_0 minimization problem:

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_0 \text{ s.t. } A\boldsymbol{z} = A\boldsymbol{x}. \tag{7}$$

A minimal requirement on ℓ_1 minimization 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., $\hat{\boldsymbol{x}} = \boldsymbol{x}$. The sufficient and necessary condition for exact ℓ_1 recovery for any $k-$sparse signal $\boldsymbol{x}$ is given by the null space condition [25]–[27]

$$\sum_{i\in S}|\boldsymbol{z}_i| < \sum_{i\notin S}|\boldsymbol{z}_i|, \forall \boldsymbol{z}\in \mathrm{Ker}(A), |S|\le k, \tag{8}$$

where $\mathrm{Ker}(A) \stackrel{\text{def}}{=} \{\boldsymbol{z} : A\boldsymbol{z} = 0\}$ is the kernel of A, and $S \subset \{1,\ldots,n\}$ is an index set. Given the sensing matrix A, the null space condition (8) imposes a bound on the level of sparsity below which exact recovery via ℓ_1 minimization is possible. Expressed in terms of $\|\cdot\|_{k,1}$, the necessary and sufficient condition becomes

$$\|\boldsymbol{z}\|_{k,1} < \frac{1}{2}\|\boldsymbol{z}\|_1,\ \forall \boldsymbol{z}\in \mathrm{Ker}(A). \tag{9}$$

The approaches in [8] and [21] for verifying the sufficient condition (9) are based on relaxing the following optimization problem in various ways:

$$\alpha_k = \max_{\boldsymbol{z}} \|\boldsymbol{z}\|_{k,1} \text{ s.t. } A\boldsymbol{z}=0, \|\boldsymbol{z}\|_1 \le 1. \tag{10}$$

Clearly, $\alpha_k < 1/2$ is necessary and sufficient for exact ℓ_1 recovery for $k-$sparse signals. Unfortunately, the direct computation of (10) for general k is difficult: it is the maximization of a norm (convex function) over a polyhedron (convex set) [28]. In [8], in a very rough sense α_1 was computed by solving n linear programs:

$$\alpha_1 = \max_{1\le i\le n}\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \|\mathbf{e}_i - A^T\boldsymbol{\lambda}\|_\infty, \tag{11}$$

where $\mathbf{e}_i$ is the ith canonical basis in $\mathbb{R}^n$. This, together with the observation that $\alpha_k < k\alpha_1$, yields an efficient algorithm to verify a sufficient condition for (9). However, in [29], we found that the primal-dual method of computing α_1 by solving n linear programs

$$\max_i \max_{\boldsymbol{z}\in\mathbb{R}^n} \boldsymbol{z}_i \text{ s.t. } A\boldsymbol{z}=0, \|\boldsymbol{z}\|_1 \le 1 \tag{12}$$

gives rise to an algorithm orders of magnitude faster. In the next section, we will see how the computation of α_1 arises naturally in the context of ℓ_∞ performance evaluation.

III. Performance Bounds on the ℓ_∞ Norms of the Recovery Errors

In this section, we derive performance bounds on the ℓ_∞ norms of the error vectors. We first establish a theorem characterizing the error vectors for the ℓ_1 recovery algorithms, whose proof is given in Appendix VIII-A

Proposition 1: Suppose $\boldsymbol{x}$ in (2) is k-sparse and the noise $\boldsymbol{w}$ satisfies $\|\boldsymbol{w}\|_\diamond \le \varepsilon$, $\|A^T\boldsymbol{w}\|_\infty \le \mu$, and $\|A^T\boldsymbol{w}\|_\infty \le \kappa\mu, \kappa \in (0,1)$, for the Basis Pursuit Denoising, the Dantzig selector, and the LASSO estimator, respectively. Define $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ as the error vector for any of the three ℓ_1 recovery algorithms (3), (4), and (5). Then we have

$$c\|\boldsymbol{h}\|_{k,1} \ge \|\boldsymbol{h}\|_1, \tag{13}$$

where $c = 2$ for the Basis Pursuit Denoising and the Dantzig selector, and $c = 2/(1-\kappa)$ for the LASSO estimator.

An immediate corollary of Proposition 1 is to bound the ℓ_1 and ℓ_2 norms of the error vector using the ℓ_∞ norm:

Corollary 1: Under the assumptions of Proposition 1, we have

$$\|\boldsymbol{h}\|_1 \le ck\|\boldsymbol{h}\|_\infty, \tag{14}$$
$$\|\boldsymbol{h}\|_2 \le \sqrt{ck}\|\boldsymbol{h}\|_\infty. \tag{15}$$

Furthermore, if $S = \mathrm{supp}(\boldsymbol{x})$ and $\beta = \min_{i\in S}|\boldsymbol{x}_i|$, then $\|\boldsymbol{h}\|_\infty < \beta/2$ implies

$$\mathrm{supp}\left(\max\left(|\hat{\boldsymbol{x}}| - \beta/2, 0\right) = \mathrm{supp}(\boldsymbol{x})\right), \tag{16}$$

i.e., a thresholding operator recovers the signal support.

We define the following quality measures for a sensing matrix A:

Definition 1: For any real number $s \in [1,n]$ and matrix $A \in \mathbb{R}^{m\times n}$, define

$$\omega_\diamond(Q,s) = \min_{\boldsymbol{z}:\|\boldsymbol{z}\|_1/\|\boldsymbol{z}\|_\infty \le s} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_\infty}, \tag{17}$$

where Q is either A or A^TA.

Now we present the error bounds on the ℓ_∞ norm of the error vectors for the Basis Pursuit Denoising, the Dantzig selector, and the LASSO estimator.

Theorem 1: Under the assumption of Proposition 1, we have

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_\infty \le \frac{2\varepsilon}{\omega_\diamond(A,2k)} \tag{18}$$

for the Basis Pursuit Denoising,

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_\infty \le \frac{2\mu}{\omega_\infty(A^TA,2k)} \tag{19}$$

for the Dantzig selector, and

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_\infty \le \frac{(1+\kappa)\mu}{\omega_\infty\left(AA^T, \frac{2k}{1-\kappa}\right)} \tag{20}$$

for the LASSO estimator.

Proof: Observe that for the Basis Pursuit Denoising

$$\begin{aligned}\|A(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_\diamond &\le \|\boldsymbol{y} - A\hat{\boldsymbol{x}}\|_\diamond + \|\boldsymbol{y} - A\boldsymbol{x}\|_\diamond \\ &\le \varepsilon + \|\boldsymbol{w}\|_\diamond \\ &\le 2\varepsilon,\end{aligned} \tag{21}$$

and similarly,

$$\left\|A^TA(\hat{\boldsymbol{x}} - \boldsymbol{x})\right\|_\infty \le 2\mu \tag{22}$$

for the Dantzig selector, and

$$\left\|A^TA(\hat{\boldsymbol{x}} - \boldsymbol{x})\right\|_\infty \le (1+\kappa)\mu \tag{23}$$

for the LASSO estimator. The conclusions of Theorem 1 follow from (14) and Definition 1. ■

In this work we design algorithms to compute $\omega_\diamond(A,s)$ and $\omega_\infty(A^TA,s)$ in polynomial time. These algorithms provide a way to numerically assess the performance of the Basis Pursuit Denoising, the Dantzig selector, and the LASSO estimator using the bounds given in Theorem 1. According to Corollary 1, the correct support can be recovered by thresholding.

In Section V, we demonstrate that the bounds in Theorem 1 are non-trivial for a large class of random sensing matrices, but they are not as tight as the RIC based bounds asymptotically. We sacrifice the tightness to get the computability. Nevertheless, numerical simulations in Section VI show that in many cases the error bounds on the ℓ_2 norms based on Corollary 1 and Theorem 1 are smaller than the RIC based bounds. We expect the bounds on the ℓ_∞ norms in Theorem 1 are even better, as we do not need the relaxation in Corollary 1.

We note that a prerequisite for the error bounds to be valid is the positiveness of $\omega_\diamond(\cdot)$. We call the validation of $\omega_\diamond(\cdot)>0$ the verification problem. Note that from Theorem 1, $\omega_\diamond(\cdot)>0$ implies the exact recovery of the true signal $\boldsymbol{x}$ in the noise-free case. Therefore, verifying $\omega_\diamond(\cdot)>0$ is equivalent to verifying a sufficient condition for exact ℓ_1 recovery.

IV. Verification and Computation of $\omega_\diamond$

In this section, we present algorithms for the verification and computation of $\omega_\diamond(\cdot)$. We will present a general algorithm and specialize it when necessary. For this purpose, we use Q to denote either A or A^TA, and use $\|\cdot\|_\diamond$ to denote a general norm.

A. Verification of $\omega_\diamond > 0$

Verifying $\omega_\diamond(Q,s)>0$ amounts to making sure $\|\boldsymbol{z}\|_1/\|\boldsymbol{z}\|_\infty > s$ for all $\boldsymbol{z}$ such that $Q\boldsymbol{z}=0$. Equivalently, we compute

$$s_* := \min\left\{\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty} : Q\boldsymbol{z}=0\right\}. \tag{24}$$

Therefore, $s<s_*$ implies the positiveness of $\omega_\diamond(Q,s)$. We rewrite the optimization problem in (24) as

$$\frac{1}{s_*} = \max_{\boldsymbol{z}} \|\boldsymbol{z}\|_\infty \text{ s.t. } Q\boldsymbol{z}=0, \|\boldsymbol{z}\|_1\le 1, \tag{25}$$

which is solved by solving n linear programs:

$$\max_i \left\{\max_{\boldsymbol{z}} \boldsymbol{z}_i \text{ s.t. } Q\boldsymbol{z}=0, \|\boldsymbol{z}\|_1\le 1\right\} \tag{26}$$

Using standard Lagrangian analysis, we show in Appendix VIII-B that the dual problem (up to a sign change) for the ith linear program in (26) is

$$\min_{\boldsymbol{\lambda}} \|\mathbf{e}_i - Q^T\boldsymbol{\lambda}\|_\infty, \tag{27}$$

where $\mathbf{e}_i$ is the ith canonical basis vector.

For general $Q=A$ or AA^T, computing s_* using (26) requires solving n linear programs. However, if A consists of row sampled from a Hadamard matrix, then the n linear programs return the same optimal value. Hence, solving one linear program suffices to compute s_*. More explicitly, we have the following corollary, whose proof is similar to the argument in Appendix B of [30] and is omitted here.

Corollary 2: If the rows of A are sampled from the rows of a Hadamard matrix, then the optimal values of

$$\max_{\boldsymbol{z}} \boldsymbol{z}_i \text{ s.t. } Q\boldsymbol{z}=0,\ \|\boldsymbol{z}\|_1\le 1 \tag{28}$$

are equal for all i.

We solve each linear program in (26) using the primal-dual algorithm expounded in Chapter 11 of [31], which gives a more efficient implementation than the one solving its dual (27) in [8]. This method is also used to implement the ℓ_1 MAGIC for sparse signal recovery [32]. Due to the equivalence of $A^TA\boldsymbol{z}=0$ and $A\boldsymbol{z}=0$, we solve (25) for $Q=A$ and avoid $Q=A^TA$. The former involves solving linear programs of smaller size. In practice, we replace A with an orthogonal basis matrix for its row space obtained from the economy-size QR decomposition of A^T.

As a dual of (27), (26) (and hence (25) and (24)) shares the same limitation as (27), namely, it verifies $\omega_\diamond>0$ only for s up to $2\sqrt{2m}$. We now reformulate Proposition 4 of [8] in our framework:

Proposition 2 [8, Proposition 4]: For any $m\times n$ matrix A with $n\ge 32m$, one has

$$s_* = \min\left\{\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty} : Q\boldsymbol{z}=0\right\} < 2\sqrt{2m}. \tag{29}$$

B. Computation of $\omega_\diamond$

Now we turn to the computation of $\omega_\diamond$. The optimization problem is the following:

$$\omega_\diamond(Q,s) = \min_{\boldsymbol{z}} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_\infty} \text{ s.t. } \frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}\le s. \tag{30}$$

Motivated by the simple fact that $\|\boldsymbol{z}\|_\infty = \max_i |\boldsymbol{z}_i|$, we define

$$\omega^i_\diamond(Q,s) = \min_{\boldsymbol{z}} \frac{\|Q\boldsymbol{z}\|_\diamond}{|\boldsymbol{z}_i|} \text{ s.t. } \frac{\|\boldsymbol{z}\|_1}{|\boldsymbol{z}_i|}\le s. \tag{31}$$

The following proposition characterize the relationship of $\omega^i_\diamond$ and $\omega_\diamond$.

Proposition 3:

$$\min_i \omega^i_\diamond(Q,s) = \omega_\diamond(Q,s). \tag{32}$$

Proof: Since $|\boldsymbol{z}_i|\le\|\boldsymbol{z}\|_\infty$, we have

$$\left\{\boldsymbol{z} : \frac{\|\boldsymbol{z}\|_1}{|\boldsymbol{z}_i|}\le s\right\} \subset \left\{\boldsymbol{z} : \frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}\le s\right\} \tag{33}$$

and

$$\frac{\|Q\boldsymbol{z}\|_\diamond}{|\boldsymbol{z}_i|} \ge \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_\infty}, \tag{34}$$

which imply that for all i

$$\begin{aligned}\omega_\diamond^i(Q,s) &= \min_{\boldsymbol{z}:\frac{\|\boldsymbol{z}\|_1}{|\boldsymbol{z}_i|}\le s} \frac{\|Q\boldsymbol{z}\|_\diamond}{|\boldsymbol{z}_i|} \\ &\ge \min_{\boldsymbol{z}:\frac{\|\boldsymbol{z}\|_1}{|\boldsymbol{z}_i|}\le s} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_\infty} \\ &\ge \min_{\boldsymbol{z}:\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}\le s} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_\infty} \\ &= \omega_\diamond(Q,s) \end{aligned} \tag{35}$$

Therefore, we get one direction of the inequality:

$$\min_i \omega_\diamond^i(Q,s) \ge \omega_\diamond(Q,s). \tag{36}$$

To prove the other direction of the inequality, suppose $\boldsymbol{z}^\star$ achieves $\omega_\diamond(Q,s)$, and $|\boldsymbol{z}_{i^\star}^\star| = \|\boldsymbol{z}^\star\|_\infty$ for index $i^\star$. Then we have the chain of equalities:

$$\omega_\diamond(Q,s) = \frac{\|Q\boldsymbol{z}^\star\|_\diamond}{\|\boldsymbol{z}^\star\|_\infty} = \frac{\|Q\boldsymbol{z}^\star\|_\diamond}{|\boldsymbol{z}_{i^\star}^\star|} \ge \omega_\diamond^{i^\star}(Q,s) \tag{37}$$

since

$$\frac{\|\boldsymbol{z}^\star\|_1}{|\boldsymbol{z}_{i^\star}^\star|} = \frac{\|\boldsymbol{z}^\star\|_1}{\|\boldsymbol{z}^\star\|_\infty} \le s. \tag{38}$$

Combination of (36) and (37) proves the proposition. ∎

The significance of Proposition 3 lies in that $\omega_\diamond^i$ can be computed in polynomial time by solving

$$\min_{\boldsymbol{z}} \|Q\boldsymbol{z}\|_\diamond \text{ s.t. } \boldsymbol{z}_i = 1, \|\boldsymbol{z}\|_1 \le s, \tag{39}$$

as long as the norm $\|\cdot\|_\diamond$ and its gradient can be computed in polynomial time. In particular, the optimization problem (39) can be reformulated as a linear program when $\diamond = 1$ and ∞, and as a quadratic program when $\diamond = 2$. By adopting an equivalent formulation of (39), we summarize the developments so far into the following theorem

Theorem 2: The quantity $\omega_\diamond(Q,s)$ is the minimum of the objective values of the following n optimizations:

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^{n-1}} \|Q_i - Q(:,-i)\boldsymbol{\lambda}\|_\diamond \text{ s.t. } \|\boldsymbol{\lambda}\|_1 \le s-1,\ i = 1,\dots,n. \tag{40}$$

Here we have used Q_i to denote the ith column of Q and $Q(:,-i)$ to denote the rest columns of Q.

Formulation (40) states that $\omega_\diamond^i$ measures how well one could approximate Q_i using a sparse combination of the rest columns of Q. We note the pairwise correlation $|Q_i^T Q_j|$, which collectively defines the mutual coherence, measures how well one could approximate Q_i using Q_j. Taking into account multiple columns of Q in the approximation allows us to get tighter error bound using $\omega_\diamond$ than that using mutual coherence.

Similar to Corollary 2, for submatrices of a Hadamard matrix, we only need to solve one optimization in (40):

Corollary 3: If the rows of A are sampled from the rows of a Hadamard matrix, then the optimal values of (39) are equal for all i.

V. Probabilistic Behavior of $\omega_\diamond(Q,s)$

In this section, we get a sense of the tightness of the error bounds based on $\omega_\diamond(Q,s)$ by deriving concentration inequalities for random sensing matrices.

We first introduce a quantity defined in [29] that is easier to analyze and can be connected to $\omega_\diamond(Q,s)$. We define the ℓ_1-constrained minimal singular value (ℓ_1-CMSV) as follows:

Definition 2: For any $s \in [1,n]$ 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

$$\rho_s(A) = \min_{\boldsymbol{z}:\|\boldsymbol{z}\|_1^2/\|\boldsymbol{z}\|_2^2\le s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}. \tag{41}$$

Despite the seeming resemblance of the definitions between $\omega_\diamond(Q,s)$, especially $\omega_2(A,s)$, and $\rho_s(A)$, the difference in the ℓ_∞ norm and the ℓ_2 norm has important implications. As shown in Theorem 2, the ℓ_∞ norm enables efficient computation of $\omega_\diamond(Q,s)$. On the other hand, the ℓ_1-CMSV yields tight performance bounds at least for a large class of random sensing matrices, as we will see in Theorem 3.

However, there are some interesting connections among these quantities, as shown in the following proposition. These connections allow us to analyze the probabilistic behavior of $\omega_\diamond(Q,s)$ using the results for $\rho_s(A)$ established in [29].

Proposition 4:

$$\sqrt{s}\sqrt{\omega_\infty(A^TA,s)} \ge \omega_2(A,s) \ge \rho_{s^2}(A). \tag{42}$$

Proof: For any $\boldsymbol{z}$ such that $\|\boldsymbol{z}\|_\infty = 1$ and $\|\boldsymbol{z}\|_1 \le s$, we have

$$\begin{aligned}\boldsymbol{z}A^TA\boldsymbol{z} &\le \sum_i |\boldsymbol{z}_i|\left|(A^TA\boldsymbol{z})_i\right| \\ &\le \left\|\boldsymbol{z}\right\|_1 \|A^TA\boldsymbol{z}\|_\infty \\ &\le s\|A^TA\boldsymbol{z}\|_\infty. \end{aligned} \tag{43}$$

Taking the minimum over $\{\boldsymbol{z} : \|\boldsymbol{z}\|_\infty = 1, \|\boldsymbol{z}\|_1 \le s\}$ yields

$$\omega_2^2(A,s) \le s\omega_\infty(A^TA,s). \tag{44}$$

Note that $\|\boldsymbol{z}\|_1/\|\boldsymbol{z}\|_\infty \le s$ implies $\|\boldsymbol{z}\|_1 \le s\|\boldsymbol{z}\|_\infty \le s\|\boldsymbol{z}\|_2$, or equivalently,

$$\{\boldsymbol{z} : \|\boldsymbol{z}\|_1/\|\boldsymbol{z}\|_\infty \le s\} \subseteq \{\boldsymbol{z} : \|\boldsymbol{z}\|_1/\|\boldsymbol{z}\|_2 \le s\}. \tag{45}$$

As a consequence, we have

$$\begin{aligned}\omega_2(A,s) &= \min_{\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}\le s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}\frac{\|\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_\infty} \\ &\ge \min_{\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty}\le s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \\ &\ge \min_{\frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_2}\le s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \\ &= \rho_{s^2}(A), \end{aligned} \tag{46}$$

where the first inequality is due to $\|\boldsymbol{z}\|_2 \ge \|\boldsymbol{z}\|_\infty$, and the second inequality is because the minimization is taken over a larger set. ∎

As a consequence of the theorem we established in [29] and include below, we derive a condition on the number of measurements to get $\omega_\diamond(Q,s)$ bounded away from zero with high probability for sensing matrices with i.i.d. subgaussian and isotropic rows. Note that a random vector $\boldsymbol{X} \in \mathbb{R}^n$ is called *isotropic and subgaussian* with constant L if $\mathbb{E}|\langle \boldsymbol{X}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\Pr(|\langle \boldsymbol{X}, \boldsymbol{u}\rangle| \geq t) \leq 2\exp(-t^2/(L\|\boldsymbol{u}\|_2))$ hold for any $\boldsymbol{u} \in \mathbb{R}^n$.

Theorem 3: [29]: 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 exist constants c_1 and c_2 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying

$$m \geq c_1 \frac{L^2 s \log n}{\epsilon^2}, \tag{47}$$

we have

$$\mathbb{E}\,|1 - \rho_s(A)| \leq \epsilon, \tag{48}$$

and

$$\mathbb{P}\left\{1 - \epsilon \leq \rho_s(A) \leq 1 + \epsilon\right\} \geq 1 - \exp(-c_2\epsilon^2 m/L^4). \tag{49}$$

A direct consequence of Proposition 4 and Theorem 3 is the following probabilistic statements about ω_2 and ω_∞.

Theorem 4: Under the assumptions and notations of Theorem 3, there exist constants c_1 and c_2 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying

$$m \geq c_1 \frac{L^2 s^2 \log n}{\epsilon^2}, \tag{50}$$

we have

$$\mathbb{E}\omega_2(A,s) \geq 1 - \epsilon, \tag{51}$$

$$\mathbb{P}\left\{\omega_2(A,s) \geq 1 - \epsilon\right\} \geq 1 - \exp(-c_2\epsilon^2 m), \tag{52}$$

and

$$\mathbb{E}\omega_\infty(A^TA,s) \geq \frac{(1-\epsilon)^2}{s}, \tag{53}$$

$$\mathbb{P}\left\{\omega_\infty(A,s) \geq \frac{(1-\epsilon)^2}{s}\right\} \geq 1 - \exp(-c_2\epsilon^2 m). \tag{54}$$

Sensing matrices with i.i.d. subgaussian and isotropic rows include the Gaussian ensemble, and the Bernoulli ensemble, as well as matrices with rows sampled according to the normalized volume measure on various convex symmetric bodies, for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [33]. In (53) and (54), the extra s in the lower bound of $\omega_\infty(A^TA,s)$ would contribute an s factor in the bounds of Theorem 1. It plays the same role as the extra $\sqrt{k}$ factor in the error bounds for the Dantzig selector and the LASSO estimator in terms of the RIC and the ℓ_1-CMSV [23], [29].

The measurement bound (50) implies that the algorithms for verifying $\omega_\diamond > 0$ and for computing $\omega_\diamond$ work for s at least up to the order $\sqrt{m/\log n}$. The order $\sqrt{m/\log n}$ is complementary to the $\sqrt{m}$ upper bound in Proposition 2.

Note that Theorem 3 implies that the following program:

$$\max_{\boldsymbol{z}} \|\boldsymbol{z}\|_2 \text{ s.t. } A\boldsymbol{z} = 0,\ \|\boldsymbol{z}\|_1 \leq 1, \tag{55}$$

verifies the sufficient condition for exact ℓ_1 recovery for s up to the order $m/\log n$, at least for subgaussian and isotropic random sensing matrices. Unfortunately, this program is NP hard and hence not tractable.

VI. Numerical Experiments

In this section, we numerically assess the performance of (40) in computing $\omega_\diamond(Q,s)$ and the tightness of bounds in Theorem 1. The numerical implementation and performance of (24) were previously reported in [29] and hence are omitted here.

When $\diamond = 1$ and ∞, the optimizations in (40) are solved using linear programs and second-order cone programs, respectively. The linear programs are implemented using the primal-dual algorithm outlined in Chapter 11 of [31]. The algorithm finds the optimal solution by solving the Karush-Kuhn-Tucker condition using linearization. The major computation is spent in solving linear systems of equations with positive definite coefficient matrices.

When $\diamond = 2$, we rewrite (40) as quadratic programs. We use the log-barrier algorithm described in Chapter 11 of [31] to solve (40). Interested readers are encouraged to refer to [32] for a concise exposition of the general primal-dual and log-barrier algorithms and implementation details for similar linear programs and second-order cone programs.

We test the algorithms on Bernoulli, Gaussian, and Hadamard matrices of different sizes. The entries of Bernoulli and Gaussian matrices are randomly generated from the classical Bernoulli distribution with equal probability and the standard Gaussian distribution, respectively. For Hadamard matrices, first a square Hadamard matrix of size n (n is a power of 2) is generated, then its rows are randomly permuted and its first m rows are taken as an $m \times n$ sensing matrix. All $m \times n$ matrices are normalized to have columns of unit length.

We compare our recovery error bounds based on $\omega_\diamond$ with those based on the RIC. Combining Corollary 1 and Theorem 1, we have for the Basis Pursuit Denoising

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2\sqrt{2k}}{\omega_2(A,2k)}\varepsilon, \tag{56}$$

and for the Dantzig selector

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2\sqrt{2k}}{\omega_\infty(A^TA,2k)}\mu. \tag{57}$$

For comparison, the two RIC bounds are

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{1+\delta_{2k}(A)}}{1-(1+\sqrt{2})\delta_{2k}(A)}\varepsilon, \tag{58}$$

for the Basis Pursuit Denoising, assuming $\delta_{2k}(A) < \sqrt{2} - 1$ [7], and

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{k}}{1-\delta_{2k}(A)-\delta_{3k}(A)}\mu, \tag{59}$$

for the Dantzig selector, assuming $\delta_{2k}(A) + \delta_{3k}(A) < 1$ [23]. Without loss of generality, we set $\varepsilon = 1$ and $\mu = 1$.

The RIC is approximated using Monte Carlo simulations. More explicitly, for $\delta_{2k}(A)$, we randomly take 1000 sub-matrices 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 $\delta_{2k}(A)$

TABLE I
COMPARISON OF THE ω_2 BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE BASIS PURSUIT DENOISING ALGORITHM FOR A BERNOULLI MATRIX WITH LEADING DIMENSION $n = 256$

	m	51	77	102	128	154	179	205
	s_*	4.6	6.1	7.4	9.6	12.1	15.2	19.3
k	k_*	2	3	3	4	6	7	9
1	ω bd	4.2	3.8	3.5	3.4	3.3	3.2	3.2
	ric bd			23.7	16.1	13.2	10.6	11.9
2	ω bd	31.4	12.2	9.0	7.4	6.5	6.0	5.6
	ric bd						72.1	192.2
3	ω bd		252.0	30.9	16.8	12.0	10.1	8.9
	ric bd							
4	ω bd				52.3	23.4	16.5	13.6
	ric bd							
5	ω bd					57.0	28.6	20.1
	ric bd							
6	ω bd					1256.6	53.6	30.8
	ric bd							
7	ω bd						161.6	50.6
	ric bd							
8	ω bd							93.1
	ric bd							
9	ω bd							258.7
	ric bd							

TABLE II
COMPARISON OF THE ω_2 BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE BASIS PURSUIT DENOISING ALGORITHM FOR A HADAMARD MATRIX WITH LEADING DIMENSION $n = 256$

	m	51	77	102	128	154	179	205
	s_*	5.4	7.1	9.1	11.4	14.0	18.4	25.3
k	k_*	2	3	4	5	6	9	12
1	ω bd	3.8	3.5	3.3	3.2	3.1	3.0	3.0
	ric bd	46.6	13.2	9.2	9.4	8.3	6.2	5.2
2	ω bd	13.7	8.4	6.7	5.9	5.4	4.9	4.6
	ric bd			46.6	24.2	15.3	8.6	7.1
3	ω bd		30.9	14.0	10.1	8.4	7.1	6.3
	ric bd				1356.6	25.4	10.3	8.8
4	ω bd			47.4	18.9	13.2	9.9	8.1
	ric bd					40.0	14.0	10.2
5	ω bd				51.5	22.6	13.8	10.3
	ric bd						18.8	11.6
6	ω bd					50.8	20.1	13.1
	ric bd						42.5	15.9
7	ω bd						31.8	16.7
	ric bd						94.2	19.7
8	ω bd						63.5	21.7
	ric bd						1000.0	24.6
9	ω bd						449.8	29.4
	ric bd							39.1
10	ω bd							42.8
	ric bd							35.6
11	ω bd							72.7
	ric bd							134.1
12	ω bd							195.1
	ric bd							

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. As a consequence, the performance bounds based on the exact RIC are *worse* than those based on the approximated RIC. Therefore, in cases where our $\omega_\diamond$ based bounds are better (tighter, smaller) than the approximated RIC bounds, they are even better than the exact RIC bounds.

In Tables I, II, and III, we compare the error bounds (56) and (58) for the Basis Pursuit Denoising algorithm. In the tables, we also include s_* computed by (25), and $k_* = \lfloor s_*/2 \rfloor$, i.e., the maximal sparsity level such that the sufficient and necessary condition (8) holds. The number of measurements is taken

TABLE III
COMPARISON OF THE ω_2 BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE BASIS PURSUIT DENOISING ALGORITHM FOR A GAUSSIAN MATRIX WITH LEADING DIMENSION $n = 256$

	m	51	77	102	128	154	179	205
	s_*	4.6	6.2	8.1	9.9	12.5	15.6	20.0
k	k_*	2	3	4	4	6	7	10
1	ω bd	4.3	3.7	3.5	3.4	3.3	3.2	3.2
	ric bd			26.0	14.2	10.0	10.9	12.1
2	ω bd	34.3	12.3	8.3	7.0	6.4	5.9	5.6
	ric bd						47.1	27.6
3	ω bd		197.4	23.4	14.5	11.6	9.8	8.9
	ric bd							
4	ω bd			1036.6	39.6	21.7	15.9	13.4
	ric bd							
5	ω bd					49.3	26.4	20.0
	ric bd							
6	ω bd					284.2	48.8	31.2
	ric bd							
7	ω bd						129.1	48.1
	ric bd							
8	ω bd							185.5
	ric bd							
9	ω bd							9640.3
	ric bd							

TABLE IV
COMPARISON OF THE ω_∞ BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE DANTZIG SELECTOR ALGORITHM FOR THE BERNOULLI MATRIX USED IN TABLE I

	m	51	77	102	128	154	179	205
	s_*	4.6	6.1	7.4	9.6	12.1	15.2	19.3
k	k_*	2	3	3	4	6	7	9
1	ω bd	6.0	5.4	4.8	4.4	4.2	4.1	4.1
	ric bd		46.3	17.4	12.1	11.2	10.3	8.6
2	ω bd	102.8	38.4	29.0	18.5	14.1	12.8	11.9
	ric bd						47.2	22.5
3	ω bd		1477.2	170.2	81.2	57.0	41.1	32.6
	ric bd							
4	ω bd				522.7	194.6	128.9	89.0
	ric bd							
5	ω bd					768.7	323.6	203.2
	ric bd							
6	ω bd					24974.0	888.7	489.0
	ric bd							
7	ω bd						3417.3	1006.9
	ric bd							
8	ω bd							2740.0
	ric bd							
9	ω bd							10196.9
	ric bd							

as $m = \lfloor \rho n \rfloor, \rho = 0.2, 0.3, \ldots, 0.8$. Note the blanks mean that the corresponding bounds are not valid. For the Bernoulli and Gaussian matrices, the RIC bounds work only for $k \leq 2$, even with $m = \lfloor 0.8n \rfloor$, while the $\omega_2(A, 2k)$ bounds work up until $k = 9$. Both bounds are better for Hadamard matrices. For example, when $m = 0.5n$, the RIC bounds are valid for $k \leq 3$, and our bounds hold for $k \leq 5$. In all cases for $n = 256$, our bounds are smaller than the RIC bounds.

We next compare the error bounds (57) and (59) for the Dantzig selector. For the Bernoulli and Gaussian matrices, our bounds work for wider ranges of (k, m) pairs and are tighter in all tested cases. For the Hadamard matrices, the RIC bounds are better, starting from $k \geq 5$ or 6. We expect that this indicates a general trend, namely, when k is relatively small, the ω based bounds are better, while when k is large, the RIC bounds are tighter. This was suggested by the probabilistic analysis of ω in Section V. The reason is that when k is relatively small, both the relaxation $\|\boldsymbol{x}\|_1 \leq 2k\|\boldsymbol{x}\|_\infty$ on the sufficient and necessary

TABLE V
COMPARISON OF THE ω_∞ BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE DANTZIG SELECTOR ALGORITHM FOR THE HADAMARD MATRIX USED IN TABLE II

	m	51	77	102	128	154	179	205
	s_*	5.2	6.9	9.1	12.1	14.4	18.3	25.2
k	k_*	2	3	4	6	7	9	12
1	ω bd	4.8	4.0	3.8	3.4	3.4	3.2	3.1
	ric bd		15.6	9.3	7.0	6.3	5.8	5.1
2	ω bd	50.9	16.2	10.1	7.1	7.0	6.1	5.3
	ric bd			45.3	16.6	13.7	10.6	8.8
3	ω bd		108.2	30.7	14.3	13.9	10.0	8.0
	ric bd			1016.4	29.9	24.9	15.8	12.5
4	ω bd			150.7	35.3	29.3	16.8	11.7
	ric bd				126.4	38.7	24.2	16.6
5	ω bd				108.5	64.2	31.4	17.3
	ric bd					187.3	30.0	22.1
6	ω bd				3168.9	171.5	59.7	25.3
	ric bd					112.0	53.1	26.8
7	ω bd					1499.5	116.3	38.8
	ric bd					411.7	71.3	34.7
8	ω bd						265.3	61.4
	ric bd						95.4	47.6
9	ω bd						2394.0	96.0
	ric bd						198.7	61.9
10	ω bd							157.4
	ric bd							82.9
11	ω bd							296.4
	ric bd							130.3
12	ω bd							898.2
	ric bd							201.2

TABLE VI
COMPARISON OF THE ω_∞ BASED BOUNDS AND THE RIC BASED BOUNDS ON THE ℓ_2 NORMS OF THE ERRORS OF THE DANTZIG SELECTOR ALGORITHM FOR THE GAUSSIAN MATRIX USED IN TABLE III

	m	51	77	102	128	154	179	205
	s_*	4.6	6.2	8.1	9.9	12.5	15.6	20.0
k	k_*	2	3	4	4	6	7	10
1	ω bd	6.5	5.1	4.8	4.3	4.2	4.0	3.9
	ric bd		30.0	18.0	14.6	9.7	9.3	9.1
2	ω bd	119.4	37.8	22.5	17.6	14.1	12.7	11.4
	ric bd					91.5	44.4	23.5
3	ω bd		1216.7	120.7	67.3	53.6	38.7	36.4
	ric bd							2546.6
4	ω bd			4515.9	318.2	168.4	115.8	109.0
	ric bd							
5	ω bd					663.6	292.4	247.8
	ric bd							
6	ω bd					5231.4	764.3	453.5
	ric bd							
7	ω bd						2646.4	1087.7
	ric bd							
8	ω bd							2450.5
	ric bd							
9	ω bd							6759.0
	ric bd							

condition (8) and the relaxation $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \sqrt{2k}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_\infty$ are sufficiently tight.

In the last set of experiments, we compare the tightness of performance bound (56) and those based on mutual coherence:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2}{1 - \mu(A)(4k-1)}\varepsilon. \tag{60}$$

for Hadamard sensing matrices with $n = 2048$ and varying m. We observe that the $\omega_2(A, 2k)$ based bound is tighter and applies to a wider range of (k, m) pairs.

VII. CONCLUSIONS

In this paper, we analyzed the performance of ℓ_1 sparse signal recovery algorithms using the ℓ_∞ norm of the errors as a performance criterion. We expressed other popular performance criteria in terms of the ℓ_∞ norm. A family of quality measures of the sensing matrices was defined using optimization procedures. We used these quality measures to derive upper bounds on the ℓ_∞ norms of the reconstruction errors for the Basis Pursuit Denoising, the Dantzig selector, and the LASSO estimator. Polynomial-time algorithms with established convergence properties were implemented to efficiently solve the optimization procedures defining the quality measures. We expect that these quality measures will be useful in comparing different sensing systems and recovery algorithms, as well as in designing optimal sensing matrices. In future work, we will use these computable performance bounds to optimally design $k-$space sample trajectories for MRI and to optimally design transmitting waveforms for compressive sensing radar.

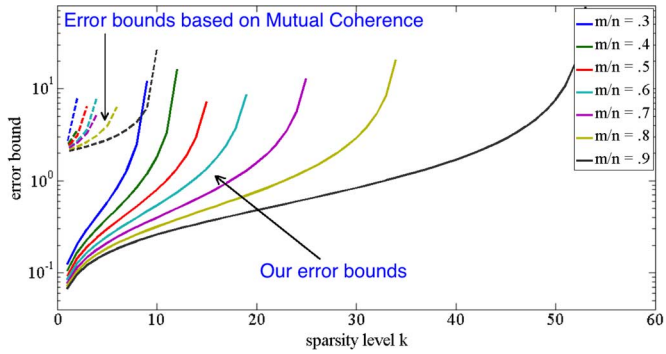

Fig. 1. $\omega_2(A, 2k)$ based bound vs. mutual coherence based bound for the Basis Pursuit for Hadamard sensing matrices with $n = 2048$.

APPENDIX
PROOFS

A. Proof of Proposition 1

Proof of Proposition 1: Suppose $S = \mathrm{supp}(\boldsymbol{x})$ 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 Basis Pursuit Denoising and the Dantzig selector. As observed by Candés in [7], the fact that $\|\hat{\boldsymbol{x}}\|_1 = \|\boldsymbol{x}+\boldsymbol{h}\|_1$ is the minimum among all $\boldsymbol{z}$s 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 Proposition 1, imply that $\|\boldsymbol{h}_{S^c}\|_1$ cannot be very large. To see this, note that

$$\begin{aligned}\|\boldsymbol{x}\|_1 &\geq \|\boldsymbol{x} + \boldsymbol{h}\|_1 \\ &= \sum_{i\in S} |\boldsymbol{x}_i + \boldsymbol{h}_i| + \sum_{i\in S^c} |\boldsymbol{x}_i + \boldsymbol{h}_i| \\ &\geq \|\boldsymbol{x}_S\|_1 - \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1 \\ &= \|\boldsymbol{x}\|_1 - \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1. \end{aligned} \tag{61}$$

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

$$2\|\boldsymbol{h}_S\|_1 \geq \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1 = \|\boldsymbol{h}\|_1. \tag{62}$$

We now turn to the LASSO estimator (5). We use the proof technique in [34] (see also [35]). Since the noise $\boldsymbol{w}$ satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\mu$ for some small $\kappa > 0$, and $\hat{\boldsymbol{x}}$ is a solution to (5), we have

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

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

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

Using the Cauchy-Swcharz type inequality, we get

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

which leads to

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

Therefore, similar to the argument in (61), we have

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

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

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

Therefore, similar to (62), we obtain

$$\begin{aligned}\frac{2}{1-\kappa}\|\boldsymbol{h}_S\|_1 &\geq \frac{1+\kappa}{1-\kappa}\|\boldsymbol{h}_S\|_1 + \frac{1-\kappa}{1-\kappa}\|\boldsymbol{h}_S\|_1\\ &\geq \frac{1+\kappa}{1-\kappa}\frac{1-\kappa}{1+\kappa}\|\boldsymbol{h}_{S^c}\|_1 + \frac{1-\kappa}{1-\kappa}\|\boldsymbol{h}_S\|_1\\ &= \|\boldsymbol{h}\|_1. \qquad (63)\end{aligned}$$

■

B. Derivation of the Dual Problem

For completeness, we derive the dual problem of the ith optimization in (26):

$$\max_{\boldsymbol{z}} \boldsymbol{z}_i \text{ s.t. } Q\boldsymbol{z} = 0,\ \|\boldsymbol{z}\|_1 \leq 1. \qquad (64)$$

Denote $\mathcal{C} = \{\boldsymbol{z} : \|\boldsymbol{z}\|_1 \leq 1\}$ a convex domain. The Lagrangian function for the primal problem (64) is

$$L(\boldsymbol{z};\boldsymbol{\lambda}) = \boldsymbol{z}_i - \boldsymbol{\lambda}^T(Q\boldsymbol{z}) = (\mathbf{e}_i - Q^T\boldsymbol{\lambda})^T\boldsymbol{z},\ \boldsymbol{z} \in \mathcal{C}$$

where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers. Therefore, the dual objective function is

$$\begin{aligned}\inf_{\boldsymbol{z}\in\mathcal{C}} L(\boldsymbol{z};\boldsymbol{\lambda}) &= \inf_{\boldsymbol{z}:\|\boldsymbol{z}\|_1\leq 1}(\mathbf{e}_i - Q^T\boldsymbol{\lambda})^T\boldsymbol{z}\\ &= -\|\mathbf{e}_i - Q^T\boldsymbol{\lambda}\|_\infty.\end{aligned}$$

Therefore, the dual problem is

$$\max_{\boldsymbol{\lambda}} -\|\mathrm{e}_i - Q^T\boldsymbol{\lambda}\|_\infty = -\min_{\boldsymbol{\lambda}}\|\mathrm{e}_i - Q^T\boldsymbol{\lambda}\|_\infty.$$

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, 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.

[3] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Apr. 2007, pp. 128–133.

[4] M. Herman and T. Strohmer, "Compressed sensing radar," in *Proc. IEEE Radar Conf.*, May 2008, pp. 1–6.

[5] 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.

[6] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.

[7] 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, pp. 589–592, 2008.

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

[9] S. A. van de Geer and P. Bhlmann, "On the conditions used to prove oracle results for the lasso," *Electron. J. Statist.*, vol. 3, pp. 1360–1392, 2009.

[10] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.

[11] M. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009.

[12] 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.

[13] 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.

[14] M. A. Sheikh, S. Sarvotham, O. Milenkovic, and R. G. Baraniuk, "DNA array decoding from nonlinear measurements by belief propagation," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Madison, WI, Aug. 2007, pp. 215–219.

[15] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[16] R. Gribonval, R. M. Figueras i Ventura, and P. Vandergheynst, "A simple test to check the optimality of a sparse signal approximation," *Signal Process.*, vol. 86, no. 3, pp. 496–510, Mar. 2006.

[17] 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.

[18] A. Tillmann and M. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, related concepts in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014.

[19] A. Bandeira, E. Dobriban, D. Mixon, and W. Sawin, "Certifying the restricted isometry property is hard," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3448–3450, Jun. 2013.

[20] A. d'Aspremont, L. E. 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.
[21] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Math. Programm.*, vol. 127, no. 1, pp. 123–144, 2011.
[22] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[23] 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.
[24] R. Tibshirani, "Regression shrinkage and selection via LASSO," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 267–288, 1996.
[25] Y. Zhang, "A simple proof for recoverability of ℓ_1-minimization: Go over or under?" CAAM Dept, Rice Univ., Houston, TX, USA, Tech. Rep., 2005.
[26] 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.
[27] D. Donoho, "High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension," Dept. of Statist., Stanford Univ., Stanford, CA, USA, Tech. Rep., 2004.
[28] H. Bodlaender, P. Gritzmann, V. Klee, and J. Leeuwen, "Computational complexity of norm-maximization," *Combinatorica*, vol. 10, pp. 203–225, 1990.
[29] G. Tang and A. Nehorai, "Performance analysis of sparse recovery based on constrained minimal singular values," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5734–5745, Dec. 2011.
[30] A. Juditsky, F. Kilinc Karzan, and A. Nemirovski, "On low rank matrix approximations with applications to synthesis problem in compressed sensing," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 3, pp. 1019–1029, 2011.
[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[32] E. Candés and J. Romberg, "ℓ_1-magic: Recovery of sparse signals via convex programming," Oct. 2005 [Online]. Available: www.acm.caltech.edu/l1magic/
[33] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geometr. Function. Anal.*, pp. 1248–1282, Nov. 2007.
[34] 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.
[35] P. Bickel, Y. Ritov, and A. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Statist.*, vol. 37, no. 4, pp. 1705–1732, 2009.

Gongguo Tang (S'09–M'11) received his Ph.D. degree in Electrical Engineering from Washington University in St. Louis in 2011, and was a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Wisconsin-Madison in 2011–2013. He is an Assistant Professor at Colorado School of Mines. His research interests are in the area of signal processing, convex optimization, information theory and statistics, 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, California. Previously, he was a faculty member at Yale University and the University of Illinois at Chicago. He is the Eugene and Martha Lohman Professor and Chair of the Preston M. Green Department of Electrical and Systems Engineering (ESE), Professor in the Department of Biomedical Engineering (by courtesy) and in the Division of Biology and Biomedical Studies (DBBS) at Washington University in St. Louis (WUSTL). He serves as Director of the Center for Sensor Signal and Information Processing at WUSTL. Under his leadership as department chair, the undergraduate enrollment has more than tripled in the last four years.

Dr. Nehorai served as Editor-in-Chief of the *IEEE Transactions on Signal Processing* from 2000 to 2002. From 2003 to 2005 he was Vice President (Publications) of the IEEE Signal Processing Society (SPS), Chair of the Publications Board, and a member of the Executive Committee of this Society. He was the founding editor of the special columns on Leadership Reflections in *IEEE Signal Processing Magazine* from 2003 to 2006.

Dr. Nehorai 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 a term lasting from 2004 to 2005. He received several best paper awards in IEEE journals and conferences. In 2001 he was named University Scholar of the University of Illinois. Dr. Nehorai was the Principal Investigator of the Multidisciplinary University Research Initiative (MURI) project titled Adaptive Waveform Diversity for Full Spectral Dominance from 2005 to 2010. He is a Fellow of the Royal Statistical Society since 1996 and Fellow of AAAS since 2012.