

# Semidefinite Programming for Computable Performance Bounds on Block-Sparsity Recovery

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

**Abstract**—In this paper, we use fixed point theory and semidefinite programming to compute the performance bounds on convex block-sparsity recovery algorithms. As a prerequisite for optimal sensing matrix design, computable performance bounds would open doors for wide applications in sensor arrays, radar, DNA microarrays, and many other areas where block-sparsity arises naturally. We define a family of quality measures for arbitrary sensing matrices as the optimal values of certain optimization problems. The reconstruction errors of convex recovery algorithms are bounded in terms of these quality measures. We demonstrate that as long as the number of measurements is relatively large, these quality measures are bounded away from zero for a large class of random sensing matrices, a result parallel to the probabilistic analysis of the block restricted isometry property. As the primary contribution of this work, we associate the quality measures with the fixed points of functions defined by a series of semidefinite programs. This relation with fixed point theory yields polynomial-time algorithms with global convergence guarantees to compute the quality measures.

**Index Terms**—Block-sparsity recovery, compressive sensing, fixed point theory, semidefinite programming.

## I. INTRODUCTION

RECOVERY of signals with low-dimensional structures, in particular, sparsity [1], block-sparsity [2], and low-rankness [3], has found numerous applications in signal sampling, inverse imaging, remote sensing, radar, sensor arrays, image processing, computer vision, and so on. Mathematically, the recovery of signals with low-dimensional structures aims to reconstruct a signal with a prescribed structure, usually from noisy linear measurements, as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^N$  is the signal to be reconstructed,  $\mathbf{y} \in \mathbb{R}^m$  is the measurement vector,  $\mathbf{A} \in \mathbb{R}^{m \times N}$  is the sensing/measurement matrix, and  $\mathbf{w} \in \mathbb{R}^m$  is the noise. For example, a sparse signal is assumed to have only a few non-zero coefficients when

represented as a linear combination of atoms from a dictionary. Similarly, for a block-sparse signal, the non-zero coefficients are assumed to cluster into blocks.

A theoretically justified way to exploit low-dimensional structure in recovering  $\mathbf{x}$  is to minimize a convex function that is known to enforce that low-dimensional structure. Examples include using the  $\ell_1$  norm to enforce sparsity, block- $\ell_1$  norm (or  $\ell_2/\ell_1$  norm) to enforce the block-sparsity, and the nuclear norm to enforce the low-rankness. The performance of these convex penalization is usually analyzed using variants of the restricted isometry property (RIP) or the coherence [2]–[5]. Upper bounds on the  $\ell_2$  norm of the errors for various recovery algorithms have been expressed in terms of the RIP or the coherence. Unfortunately, it is extremely difficult to verify that the RIP of a specific sensing matrix satisfies the conditions for the bounds to be valid, and even more difficult to directly compute the RIP constant itself. Actually, the only known sensing matrices with nice RIPs are certain types of random matrices [6]. The coherence of sensing matrices are easy to compute, but they typically offer weak error bounds [5].

In this paper, we investigate the recovery performance for block-sparse signals. Block-sparsity arises naturally in applications such as sensor arrays [7], radar [8], multi-band signals [9], and DNA microarrays [10]. A particular area that motivates this work is the application of block-sparse signal recovery in radar systems. The signals in radar applications are usually sparse because there are only a few targets to be estimated among many possibilities. However, a single target manifests itself simultaneously in the sensor domain, the frequency domain, the temporal domain, and the reflection-path domain. As a consequence, the underlying signal would be block-sparse when the radar system observes the targets from several of these domains [8], [11].

The aforementioned applications require a computable performance analysis of block-sparsity recovery. While it is reasonable to use a random sensing matrix for signal sampling, the sensing matrices in other applications are far from random and depend on the underlying measurement devices and the physical processes that generate the observations. Due to the computational challenges associated with the RIP, it is necessary to seek computationally more amenable quality measures of the sensing matrices. Computable performance bounds would open doors for wide applications. They would provide a means to pre-determine the performance of the sensing system before its implementation and the taking of measurements. In addition, in applications where we have the flexibility to choose the sensing matrix, computable performance bounds would form a basis for optimal sensing matrix design.

We preview our contributions. First of all, we define a family of quality measures of the sensing matrix, and use them to derive

Manuscript received November 12, 2014; revised September 20, 2015; accepted November 12, 2015. Date of publication May 12, 2016; date of current version July 21, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Joakim Jaldén. This work was supported by AFOSR Grant FA9550-11-1-0210, by NSF Grants CCF-1464205, CCF-1014908, and CCF-0963742, and by ONR Grant N000141310050. This work appeared in part in Computable Performance Analysis of Block-Sparsity Recovery in the *Proceedings of the 4th IEEE International Workshop on Computational Advances in Multisensor Adaptive Processing (CAMSAP)*, 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).

Digital Object Identifier 10.1109/TSP.2016.2568166

performance bounds on the block- $\ell_\infty$  norm of the recovery error vector. Performance bounds using other norms are expressed using the block- $\ell_\infty$  norm. Secondly, we develop a fixed point iteration framework to design algorithms that compute lower bounds on the quality measures for arbitrary sensing matrices. Each fixed point iteration solves a series of semidefinite programs. The fixed point iteration framework also demonstrates the algorithms' convergence to the global optimum from any initial point. As by-products, we obtain algorithms to verify the sufficient condition guaranteeing exact block-sparsity recovery via block- $\ell_1$  minimization. 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 that of the block RIP for random matrices. The sampling complexities derived from our quality measures are higher than those based on the block RIP, even though numerical results in Section VI show that our performance bounds are smaller (tighter) than the RIP-based bounds for the performed experiments. Preliminary results of the present work were given in its conference version [12], which omitted all the proofs.

This work extends verifiable and computable performance analysis from sparse signal cases [6], [13]–[15] to block-sparse signal cases. There are several technical challenges that are unique to the block-sparse case. In the sparse setting, the optimization subproblems are solved exactly by linear programming or second-order cone programming. However, in the block-sparse setting, the associated subproblems are not readily solvable and we need to develop semidefinite relaxations to compute an upper bound on the optimal values of the subproblems. The bounding technique complicates the situation as it is not clear how to obtain bounds on the quality measures from these subproblem bounds. We develop a systematic fixed point theory to address this problem.

The rest of the paper is organized as follows. In Section II, we present the measurement model, three convex relaxation algorithms, and the condition for exact block- $\ell_1$  recovery. In Section III, we derive performance bounds on the block- $\ell_\infty$  norms of the recovery errors for several convex relaxation algorithms. Section IV is devoted to the probabilistic analysis of our block- $\ell_\infty$  performance measures. In Section V, we design algorithms to verify a sufficient condition for exact block- $\ell_1$  recovery in the noise-free case, and to compute the quality measures of arbitrary sensing matrices through fixed point iteration, bisection search, and semidefinite programming. 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 present three block-sparsity recovery algorithms.

For any vector  $\mathbf{x} \in \mathbb{R}^{np}$ , we partition the vector into  $p$  blocks, each of length  $n$ . More precisely, we have  $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_p^T]^T$ , with the  $i$ th block  $\mathbf{x}_i \in \mathbb{R}^n$ . The block- $\ell_q$  norms for  $1 \leq q \leq \infty$

associated with this block-sparse structure are defined as

$$\|\mathbf{x}\|_{bq} = \left( \sum_{i=1}^p \|\mathbf{x}_i\|_2^q \right)^{1/q}, \quad 1 \leq q < \infty \quad (2)$$

and

$$\|\mathbf{x}\|_{b\infty} = \max_{1 \leq i \leq p} \|\mathbf{x}_i\|_2, \quad q = \infty. \quad (3)$$

The canonical inner product in  $\mathbb{R}^{np}$  is denoted by  $\langle \cdot, \cdot \rangle$ , and the  $\ell_2$  (or Euclidean) norm is  $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ . Obviously, the block- $\ell_2$  norm is the same as the ordinary  $\ell_2$  norm. We use  $\|\cdot\|_\diamond$  to denote a general norm.

The block support of  $\mathbf{x} \in \mathbb{R}^{np}$ ,  $\text{bsupp}(\mathbf{x}) = \{i : \|\mathbf{x}_i\|_2 \neq 0\}$ , is the index set of the non-zero blocks of  $\mathbf{x}$ . The size of the block support, denoted by the block- $\ell_0$  "norm"  $\|\mathbf{x}\|_{b0}$ , is the block-sparsity level of  $\mathbf{x}$ . Signals of block-sparsity level of at most  $k$  are called  $k$ -block-sparse signals. If  $S \subset \{1, \dots, p\}$  is an index set, then  $|S|$  is the cardinality of  $S$ , and  $\mathbf{x}_S \in \mathbb{R}^{n|S|}$  is the vector formed by the blocks of  $\mathbf{x}$  with indices in  $S$ .

We use  $\mathbf{e}_i$ ,  $\mathbf{0}$ ,  $\mathbf{O}$ ,  $\mathbf{1}$ , and  $\mathbf{I}_N$  to denote respectively the  $i$ th canonical basis vector, the zero vector, the zero matrix, the vector with all ones, and the identity matrix of size  $N \times N$ .

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

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad (4)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times np}$  is the measurement/sensing matrix,  $\mathbf{y}$  is the measurement vector, and  $\mathbf{w}$  is noise. A very special block-sparse model is when  $\mathbf{x}$  is a complex signal, such as the models in sensor arrays and radar applications. Note that the computable performance analysis developed in [13], [16] for the real variables can not apply to the complex case.

Many algorithms for sparse recovery have been extended to recover the block-sparse signal  $\mathbf{x}$  from  $\mathbf{y}$  by exploiting the block-sparsity of  $\mathbf{x}$ . We focus on three algorithms based on block- $\ell_1$  minimization: the Block-Sparse Basis Pursuit (BS-BP) [2], the Block-Sparse Dantzig selector (BS-DS) [17], and the Block-Sparse LASSO estimator (BS-LASSO) [18].

$$\text{BS-BP} : \min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{z}\|_{b1} \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \leq \varepsilon \quad (5)$$

$$\text{BS-DS} : \min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{z}\|_{b1} \quad \text{s.t.} \quad \|\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{z})\|_{b\infty} \leq \mu \quad (6)$$

$$\text{BS-LASSO} : \min_{\mathbf{z} \in \mathbb{R}^{np}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \mu \|\mathbf{z}\|_{b1}. \quad (7)$$

Here  $\mu$  is a tuning parameter, and  $\varepsilon$  measures the noise level.

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

$$\min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{z}\|_{b1} \quad \text{s.t.} \quad \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}, \quad (8)$$

which is the block- $\ell_1$  relaxation of the block- $\ell_0$  problem:

$$\min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{z}\|_{b0} \quad \text{s.t.} \quad \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \quad (9)$$

A minimal requirement on the block- $\ell_1$  minimization algorithms is the *uniqueness and exactness* of the solution

$\hat{\mathbf{x}} \stackrel{\text{def}}{=} \operatorname{argmin}_{\mathbf{z}: A\mathbf{z} = A\mathbf{x}} \|\mathbf{x}\|_{b1}$ , i.e.,  $\hat{\mathbf{x}} = \mathbf{x}$ . When the true signal  $\mathbf{x}$  is  $k$ -block-sparse, the sufficient and necessary condition for exact block- $\ell_1$  recovery is [19]

$$\sum_{i \in S} \|\mathbf{z}_i\|_2 < \sum_{i \notin S} \|\mathbf{z}_i\|_2, \forall \mathbf{z} \in \operatorname{Ker}(A), |S| \leq k, \quad (10)$$

where  $\operatorname{Ker}(A) \stackrel{\text{def}}{=} \{\mathbf{z} : A\mathbf{z} = 0\}$  is the kernel of  $A$ ,  $S \subset \{1, \dots, p\}$  is an index set, and  $\mathbf{z}_i$  is the  $i$ th size- $n$  block of  $\mathbf{z}$ .

### III. PERFORMANCE BOUNDS ON THE BLOCK- $\ell_\infty$ NORMS OF THE RECOVERY ERRORS

In this section, we derive performance bounds on the block- $\ell_\infty$  norms of the error vectors. We first establish a proposition characterizing the error vectors of the block- $\ell_1$  recovery algorithms, whose proof is given in Appendix VIII-A.

*Proposition 1:* Suppose  $\mathbf{x} \in \mathbb{R}^{np}$  in (4) is  $k$ -block-sparse and the noise  $\mathbf{w}$  satisfies  $\|\mathbf{w}\|_2 \leq \varepsilon$ ,  $\|A^T \mathbf{w}\|_{b\infty} \leq \mu$ , and  $\|A^T \mathbf{w}\|_\infty \leq \kappa\mu$ ,  $\kappa \in (0, 1)$ , for the BS-BP, the BS-DS, and the BS-LASSO, respectively. Define  $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}$  as the error vector for any of the three block- $\ell_1$  recovery algorithms (5)–(7). Then we have

$$\|\mathbf{h}_S\|_{b1} \geq \|\mathbf{h}\|_{b1}/c, \quad (11)$$

where  $S = \operatorname{bsupp}(\mathbf{x})$ ,  $c = 2$  for the BS-BP and the BS-DS, and  $c = 2/(1 - \kappa)$  for the BS-LASSO.

An immediate corollary of Proposition 1 is to bound the block- $\ell_1$  and  $\ell_2$  norms of the error vectors, using the block- $\ell_\infty$  norm. The proof is given in Appendix VIII-B.

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

$$\|\mathbf{h}\|_{b1} \leq ck \|\mathbf{h}\|_{b\infty}, \text{ and} \quad (12)$$

$$\|\mathbf{h}\|_{l2} \leq \sqrt{ck} \|\mathbf{h}\|_{b\infty}. \quad (13)$$

Furthermore, if  $S = \operatorname{bsupp}(\mathbf{x})$  and  $\beta = \min_{i \in S} \|\mathbf{x}_i\|_2$ , then  $\|\mathbf{h}\|_{b\infty} < \beta/2$  implies

$$\{i : \|\hat{\mathbf{x}}_i\|_2 > \beta/2\} = \operatorname{bsupp}(\mathbf{x}), \quad (14)$$

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

We need the following notation, which measures the stability of the sensing matrix when viewed as a linear mapping with sparse vectors as inputs. It is the key quantity for our computable performance bounds.

*Definition 1:* For  $s \in [1, p]$  and matrix  $A \in \mathbb{R}^{m \times np}$ , define

$$\omega_\diamond(Q, s) = \min_{\mathbf{z}: \|\mathbf{z}\|_{b1}/\|\mathbf{z}\|_{b\infty} \leq s} \frac{\|Q\mathbf{z}\|_\diamond}{\|\mathbf{z}\|_{b\infty}}, \quad (15)$$

where  $Q$  is either  $A$  or  $A^T A$ .

Now we present the error bounds on the block- $\ell_\infty$  norm of the error vectors for the BS-BP, the BS-DS, and the BS-LASSO, whose proof is given in Appendix VIII-C.

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

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_{b\infty} \leq \frac{2\varepsilon}{\omega_2(A, 2k)} \quad (16)$$

for the BS-BP,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_{b\infty} \leq \frac{2\mu}{\omega_{b\infty}(A^T A, 2k)} \quad (17)$$

for the BS-DS, and

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_{b\infty} \leq \frac{(1 + \kappa)\mu}{\omega_{b\infty}(A^T A, 2k/(1 - \kappa))} \quad (18)$$

for the BS-LASSO.

A consequence of Theorem 1 and Corollary 1 is the error bound on the  $\ell_2$  norm:

*Corollary 2:* Under the assumption of Proposition 1, the  $\ell_2$  norms of the recovery errors are bounded as

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2\sqrt{2k}\varepsilon}{\omega_2(A, 2k)} \quad (19)$$

for the BS-BP,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2\sqrt{2k}\mu}{\omega_{b\infty}(A^T A, 2k)} \quad (20)$$

for the BS-DS, and

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \sqrt{\frac{2k}{1 - \kappa}} \frac{(1 + \kappa)\mu}{\omega_{b\infty}(A^T A, 2k/(1 - \kappa))} \quad (21)$$

for the BS-LASSO.

One of the contributions of this work is the design of algorithms to compute  $\omega_\diamond(A, s)$  and  $\omega_{b\infty}(A^T A, s)$ . The algorithms provide a way to numerically assess the performance of the BS-BP, the BS-DS, and the BS-LASSO according to the bounds given in Theorem 1 and Corollary 2. According to Corollary 1, the correct recovery of signal block-support is also guaranteed by reducing the block- $\ell_\infty$  norm to some threshold. In Section IV, we demonstrate that the bounds in Theorem 1 are non-trivial for a large class of random sensing matrices, as long as  $m$  is relatively large. Numerical simulations in Section VI show that in many cases the error bounds of Corollary 2 are tighter than the block RIP based bounds, even though a definite comparison of these bounds requires more extensive investigations. We next establish some results on the probabilistic behavior of  $\omega_\diamond(Q, s)$ .

### IV. PROBABILISTIC BEHAVIOR OF $\omega_\diamond(Q, s)$

In this section, we analyze how good are the performance bounds in Theorem 1 for random sensing matrices. For this purpose, we define the block  $\ell_1$ -constrained minimal singular value (block  $\ell_1$ -CMSV), which is an extension of the  $\ell_1$ -CMSV concept in the sparse setting [16]:

*Definition 2:* For any  $s \in [1, p]$  and matrix  $A \in \mathbb{R}^{m \times np}$ , define the block  $\ell_1$ -constrained minimal singular value (abbreviated as block  $\ell_1$ -CMSV) of  $A$  by

$$\rho_s(A) = \min_{\mathbf{z}: \|\mathbf{z}\|_{b1}^2/\|\mathbf{z}\|_2^2 \leq s} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2}. \quad (22)$$

The most important difference between  $\rho_s(A)$  and  $\omega_\diamond(Q, s)$  is the replacement of  $\|\cdot\|_{b\infty}$  with  $\|\cdot\|_2$  in the denominators of the fractional constraint and the objective function. The Euclidean norm  $\|\cdot\|_2$  is more amenable to probabilistic analysis. The connections between  $\rho_s(A)$ ,  $\omega_2(A, s)$ , and

$\omega_{\text{b}\infty}(A^T A, s)$ , established in the following lemma, allow us to analyze the probabilistic behaviors of  $\omega_\diamond(Q, s)$  using the results for  $\rho_s(A)$  which we are going to establish later. The proof is given in Appendix VIII-D.

*Lemma 1:*

$$\sqrt{s}\sqrt{\omega_{\text{b}\infty}(A^T A, s)} \geq \omega_2(A, s) \geq \rho_{s^2}(A). \quad (23)$$

Next we derive a condition on the number of measurements needed for  $\rho_s(A)$  to be bounded away from zero with high probability for sensing matrices with *i.i.d.* subgaussian and isotropic rows. A random vector  $\mathbf{X} \in \mathbb{R}^{np}$  is called *isotropic and subgaussian* with constant  $L$  if  $\mathbb{E}|\langle \mathbf{X}, \mathbf{u} \rangle|^2 = \|\mathbf{u}\|_2^2$  and  $\mathbb{P}(|\langle \mathbf{X}, \mathbf{u} \rangle| \geq t) \leq 2 \exp(-t^2 / (L \|\mathbf{u}\|_2))$  hold for any  $\mathbf{u} \in \mathbb{R}^{np}$ .

*Theorem 2:* Let the rows of  $\sqrt{m}A$  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

$$m \geq c_1 \frac{L^2 (sn + s \log p)}{\epsilon^2}, \quad (24)$$

we have

$$\mathbb{E}\rho_s(A) \geq 1 - \epsilon, \quad (25)$$

and

$$\mathbb{P}\{\rho_s(A) \geq 1 - \epsilon\} \geq 1 - \exp(-c_2 \epsilon^2 m / L^4). \quad (26)$$

For a proof, see Appendix VIII-E.

Using  $\rho_s(A)$ , we could equally develop bounds similar to those of Theorem 1 on the  $\ell_2$  norm of the error vectors. For example, the error bound for the BS-BP would look like

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2\epsilon}{\rho_{2k}(A)}. \quad (27)$$

The conclusion of Theorem 2, combined with the previous equation, implies that we can recover a block sparse signal stably using BS-BP with high probability if the sensing matrix is subgaussian and isotropic and  $m \geq c(kn + k \log p)/\epsilon^2$ . If we do not consider the block structure in the signal, we would need  $m \geq c(kn \log(np))/\epsilon^2$  measurements, as the sparsity level is  $kn$  [16]. Therefore, the prior information regarding the block structure greatly reduces the number of measurements necessary to recover the signal. The lower bound on  $m$  is essentially the same as the one given by the block RIP (See ([2], Proposition 4)).

*Theorem 3:* Under the assumptions and notations of Theorem 2, there exist constants  $c_1$  and  $c_2$  such that for any  $\epsilon > 0$  and  $m \geq 1$  satisfying

$$m \geq c_1 \frac{L^2 (s^2 n + s^2 \log p)}{\epsilon^2}, \quad (28)$$

we have

$$\mathbb{E}\omega_2(A, s) \geq 1 - \epsilon, \text{ and} \quad (29)$$

$$\mathbb{P}\{\omega_2(A, s) \geq 1 - \epsilon\} \geq 1 - \exp(-c_2 \epsilon^2 m / L^4), \quad (30)$$

and

$$\mathbb{E}\omega_{\text{b}\infty}(A^T A, s) \geq \frac{(1 - \epsilon)^2}{s}, \text{ and} \quad (31)$$

$$\mathbb{P}\left\{\omega_{\text{b}\infty}(A^T A, s) \geq \frac{(1 - \epsilon)^2}{s}\right\} \geq 1 - \exp(-c_2 \epsilon^2 m / L^4). \quad (32)$$

Equation (28) and Theorem 1 imply that for exact signal recovery in the noise-free case, we need  $O(s^2(n + \log p))$  measurements for random sensing matrices. The extra  $s$  suggests that the  $\omega_\diamond$  based approach to verify exact recovery is not as good as the one based on  $\rho_s$ . However,  $\omega_\diamond$  is computationally more amenable, as we are going to see in Section V. The measurement bound (28) also implies that the algorithms for verifying  $\omega_\diamond > 0$  and for computing  $\omega_\diamond$  work for  $s$  at least up to the order  $\sqrt{m/(n + \log p)}$ .

Finally, we comment that sensing matrices with *i.i.d.* subgaussian and isotropic rows include the Gaussian ensemble, the Bernoulli ensemble, and normalized volume measure on various convex symmetric bodies, for example, the unit balls of  $\ell_q^{np}$  for  $2 \leq q \leq \infty$  [20].

## V. VERIFICATION AND COMPUTATION OF $\omega_\diamond$

We now consider the computation of  $\omega_\diamond(\cdot)$ . We will present a general algorithm and make it specific when necessary. For this purpose, we use  $Q$  to denote either  $A$  or  $A^T A$ .

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

A prerequisite for the bounds in Theorem 1 to be valid is the positivity 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 exact recovery of the true signal  $\mathbf{x}$  in the noise-free case. Therefore, verifying  $\omega_\diamond(\cdot) > 0$  is equivalent to verifying a sufficient condition for exact block- $\ell_1$  recovery.

Verifying  $\omega_\diamond(Q, s) > 0$  amounts to making sure  $\|\mathbf{z}\|_{\text{b}1} / \|\mathbf{z}\|_{\text{b}\infty} \leq s$  for all  $\mathbf{z}$  in the null space of  $Q$ . Therefore, we compute

$$s^* = \min_{\mathbf{z}} \frac{\|\mathbf{z}\|_{\text{b}1}}{\|\mathbf{z}\|_{\text{b}\infty}} \text{ s.t. } Q\mathbf{z} = 0. \quad (33)$$

Then, when  $s < s^*$ , we have  $\omega_\diamond(Q, s) > 0$ . The following theorem presents an optimization procedure that computes a lower bound on  $s^*$ .

*Proposition 2:* Denote by  $\text{opt}$  the optimal value of the following optimization

$$\max_i \min_{P_i} \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2. \quad (34)$$

Its inverse  $s_* = 1/\text{opt}$  is a lower bound on  $s^*$ . Here  $P$  is a matrix variable of the same size as  $Q$ ,  $\delta_{ij} = 1$  for  $i = j$  and 0 otherwise, and  $P = [P_1, \dots, P_p]$ ,  $Q = [Q_1, \dots, Q_p]$  with  $P_i$  and  $Q_j$  having  $n$  columns each.

The proof is given in Appendix VIII-F.

Because  $s_* < s^*$ , the condition  $s < s_*$  is a sufficient condition for  $\omega_\diamond > 0$  and for the uniqueness and exactness of block-sparse



recovery in the noise-free case. To get  $s_*$ , for each  $i$ , we need to solve

$$\min_{P_i} \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2. \quad (35)$$

A semidefinite program equivalent to (35) is given as follows:

$$\min_{P_i, t} t \text{ s.t. } \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \leq t, j = 1, \dots, p, \quad (36)$$

$$\Leftrightarrow \min_{P_i, t} t \text{ s.t. } \begin{bmatrix} t \mathbf{I}_n & \delta_{ij} \mathbf{I}_n - P_i^T Q_j \\ \delta_{ij} \mathbf{I}_n - Q_j^T P_i & t \mathbf{I}_n \end{bmatrix} \succeq 0, \quad (37)$$

$$j = 1, \dots, p.$$

Small instances of (36) and (37) can be solved using CVX [21]. However, for medium to large scale problems, it is beneficial to use first-order techniques to solve (35) directly. We observe that  $\max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2$  can be expressed as the largest eigenvalue of a block-diagonal matrix. The smoothing technique for semidefinite optimization developed in [22] can be used to minimize the largest eigenvalue with respect to  $P_i$ . We leave implementations issues in future work.

Due to the equivalence of  $A^T A z = 0$  and  $A z = 0$ , we always solve (33) for  $Q = A$  and avoid  $Q = A^T A$ . The former apparently involves solving semidefinite programs of smaller size. In practice, we usually replace  $A$  with the matrix with orthogonal rows obtained from the economy-size QR decomposition of  $A^T$ .

### B. Fixed Point Theory for Computing $\omega_\diamond(\cdot)$

We present a general fixed point procedure to compute  $\omega_\diamond$ . Recall that the optimization problem defining  $\omega_\diamond$  is as follows:

$$\omega_\diamond(Q, s) = \min_z \frac{\|Qz\|_\diamond}{\|z\|_{b\infty}} \text{ s.t. } \frac{\|z\|_{b1}}{\|z\|_{b\infty}} \leq s. \quad (38)$$

Motivated by the fact that  $\|z\|_{b\infty} = \max_{1 \leq i \leq p} \|z_i\|_2$ , we defined a set of closely related quantities

$$\omega_\diamond^i(Q, s) = \min_z \frac{\|Qz\|_\diamond}{\|z_i\|_2} \text{ s.t. } \frac{\|z\|_{b1}}{\|z_i\|_2} \leq s. \quad (39)$$

The following proposition characterizes the relationship of  $\omega_\diamond^i$  and  $\omega_\diamond$ .

*Proposition 3:*

$$\min_i \omega_\diamond^i(Q, s) = \omega_\diamond(Q, s). \quad (40)$$

*Proof:* Since  $\|z_i\|_2 \leq \|z\|_{b\infty}$ , we have

$$\left\{ z : \frac{\|z\|_{b1}}{\|z_i\|_2} \leq s \right\} \subset \left\{ z : \frac{\|z\|_{b1}}{\|z\|_{b\infty}} \leq s \right\} \quad (41)$$

and

$$\frac{\|Qz\|_\diamond}{\|z_i\|_2} \geq \frac{\|Qz\|_\diamond}{\|z\|_{b\infty}}, \quad (42)$$

which imply that for all  $i$

$$\begin{aligned} \omega_\diamond^i(Q, s) &= \min_{z: \frac{\|z\|_{b1}}{\|z_i\|_2} \leq s} \frac{\|Qz\|_\diamond}{\|z_i\|_2} \geq \min_{z: \frac{\|z\|_{b1}}{\|z\|_{b\infty}} \leq s} \frac{\|Qz\|_\diamond}{\|z\|_{b\infty}} \\ &\geq \min_{z: \frac{\|z\|_{b1}}{\|z\|_{b\infty}} \leq s} \frac{\|Qz\|_\diamond}{\|z\|_{b\infty}} = \omega_\diamond(Q, s) \end{aligned} \quad (43)$$

Therefore, we get one direction of the inequality:

$$\min_i \omega_\diamond^i(Q, s) \geq \omega_\diamond(Q, s). \quad (44)$$

To prove the other direction of the inequality, suppose  $z^*$  achieves  $\omega_\diamond(Q, s)$ , and  $\|z_{i^*}^*\|_2 = \|z^*\|_{b\infty}$  for index  $i^*$ . Then we have the chain of equalities:

$$\omega_\diamond(Q, s) = \frac{\|Qz^*\|_\diamond}{\|z^*\|_{b\infty}} = \frac{\|Qz^*\|_\diamond}{\|z_{i^*}^*\|_2} \geq \omega_\diamond^{i^*}(Q, s) \quad (45)$$

since

$$\frac{\|z^*\|_{b1}}{\|z_{i^*}^*\|_2} = \frac{\|z^*\|_{b1}}{\|z^*\|_{b\infty}} \leq s. \quad (46)$$

Combination of (44) and (45) proves the proposition.  $\blacksquare$

We have transformed the computation of  $\omega_\diamond(Q, s)$  into the computation of  $\omega_\diamond^i(Q, s)$ . Unfortunately, it is still difficult to compute  $\omega_\diamond^i(Q, s)$  exactly in polynomial time.

In the following, we first characterize  $\omega_\diamond^i$  as the fixed point of a scalar function. For any  $s \in (1, s^*)$  and  $i \in \{1, \dots, p\}$ , define a function over  $[0, \infty)$  parameterized by  $s$

$$f_{s,i}(\eta) = \max_z \{ \|z_i\|_2 : \|Qz\|_\diamond \leq 1, \|z\|_{b1} \leq s\eta \}. \quad (47)$$

We then have the following proposition, whose proof is given in Appendix VIII-G.

*Proposition 4:*  $f_{s,i}(\eta)$  has a unique positive fixed point  $\eta_i^* = f_{s,i}(\eta_i^*)$  that is equal to  $1/\omega_\diamond^i(Q, s)$ .

Unfortunately, except when  $n = 1$  and the signal is real, i.e., the real sparse case, it is not easy to compute  $f_{s,i}(\eta)$  according to (47). In the following proposition, we present a relaxation of the problem

$$\max_z \|z_i\|_2 \text{ s.t. } \|Qz\|_\diamond \leq 1, \|z\|_{b1} \leq s\eta \quad (48)$$

to compute an upper bound on  $f_{s,i}(\eta)$ . This proof is similar to that of Proposition 2 and is given in Appendix VIII-H.

*Proposition 5:* When  $Q = A$  and  $\diamond = 2$ , we have

$$f_{s,i}(\eta) \leq \min_{P_i} \max_j s\eta \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 + \|P_i\|_2, \quad (49)$$

and when  $Q = A^T A$  and  $\diamond = b\infty$ , we have

$$f_{s,i}(\eta) \leq \min_{P_i} \max_j s\eta \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 + \sum_{l=1}^p \|P_i^l\|_2. \quad (50)$$

Here  $P_i$  is a matrix with the same number of rows as  $Q$  and  $n$  columns,  $Q_j$  is the submatrix of  $Q$  formed by the  $(j-1)n+1$ th to  $jn$ th columns, and  $P_i^l$  is the submatrix of  $P_i$  formed by the  $(l-1)n+1$ th to  $ln$ th rows.

For each  $i = 1, \dots, p$ , the optimization problem

$$\min_{P_i} \max_j s\eta \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 + \|P_i\|_2 \quad (51)$$

can be solved using semidefinite programming:

$$\begin{aligned}
& \min_{P_i, t_0, t_1} s\eta t_0 + t_1 \\
& \text{s.t. } \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \leq t_0, j = 1, \dots, p, \|P_i\|_2 \leq t_1, \\
& \Leftrightarrow \min_{P_i, t_0, t_1} s\eta t_0 + t_1 \text{ s.t.} \\
& \begin{bmatrix} t_0 \mathbf{I}_n & \delta_{ij} \mathbf{I}_n - P_i^T Q_j \\ \delta_{ij} \mathbf{I}_n - Q_j^T P_i & t_0 \mathbf{I}_n \end{bmatrix} \succeq 0, j = 1, \dots, p, \\
& \begin{bmatrix} t_1 \mathbf{I}_m & P_i \\ P_i^T & t_1 \mathbf{I}_n \end{bmatrix} \succeq 0. \quad (52)
\end{aligned}$$

Similarly, the optimization problem

$$\min_{P_i} \max_j s\eta \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 + \sum_{l=1}^p \|P_i^l\|_2 \quad (53)$$

can be solved by the following semidefinite program:

$$\begin{aligned}
& \min_{P_i, t_0, t_1, \dots, t_p} s\eta t_0 + \sum_{l=1}^p t_l \\
& \text{s.t. } \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \leq t_0, j = 1, \dots, p, \\
& \|P_i^l\|_2 \leq t_l, l = 1, \dots, p, \\
& \Leftrightarrow \min_{P_i, t_0, t_1} s\eta t_0 + \sum_{l=1}^p t_l \text{ s.t.} \\
& \begin{bmatrix} t_0 \mathbf{I}_n & \delta_{ij} \mathbf{I}_n - P_i^T Q_j \\ \delta_{ij} \mathbf{I}_n - Q_j^T P_i & t_0 \mathbf{I}_n \end{bmatrix} \succeq 0, j = 1, \dots, p, \\
& \begin{bmatrix} t_l \mathbf{I}_n & P_i^l \\ P_i^{lT} & t_l \mathbf{I}_n \end{bmatrix} \succeq 0, l = 1, \dots, p. \quad (54)
\end{aligned}$$

These semidefinite programs can also be solved using first-order techniques.

### C. Fixed Point Theory for Computing a Lower Bound on $\omega_\circ$

Proposition 5 provides ways to compute an upper bound on  $f_{s,i}(\eta)$  for fixed  $\eta$ . It turns out we could use it to compute an upper bound on the positive fixed point of  $f_{s,i}(\eta)$ , or  $1/\omega_\circ^i(Q, s)$ .

We need a few definitions. When  $Q = A$  and  $\|\cdot\|_\circ = \|\cdot\|_2$ , define functions  $g_{s,i}(\eta)$  over  $[0, \infty)$  parameterized by  $s \in (1, s_*)$  as

$$g_{s,i}(\eta) = \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2. \quad (55)$$

Similarly, when  $Q = A^T A$  and  $\|\cdot\|_\circ = \|\cdot\|_{b\infty}$ , we define:

$$h_{s,i}(\eta) = \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) + \sum_{l=1}^p \|P_i^l\|_2 \quad (56)$$

The following proposition, whose proof is given in Appendix VIII-I, lists some properties of  $g_{s,i}(\eta)$  and  $h_{s,i}(\eta)$ .

**Proposition 6:** The functions  $g_{s,i}(\eta)$  have the following properties. The same is true for  $h_{s,i}(\eta)$ .

- 1)  $g_{s,i}(\eta)$  is continuous, strictly increasing, and concave as a function of  $\eta$ .

- 2)  $g_{s,i}(0) = 0, g_{s,i}(\eta) \geq s\eta > \eta$  for sufficiently small  $\eta > 0$ , and there exists  $\rho < 1$  such that  $g_{s,i}(\eta) < \rho\eta$  for sufficiently large  $\eta$ .
- 3)  $g_{s,i}$  has a unique positive fixed point  $\bar{\eta}_i^* = g_{s,i}(\bar{\eta}_i^*)$ .
- 4) For  $\eta \in (0, \bar{\eta}_i^*)$ , we have  $g_{s,i}(\eta) > \eta$ , and for  $\eta \in (\bar{\eta}_i^*, \infty)$ , we have  $g_{s,i}(\eta) < \eta$ .
- 5) For any  $\epsilon > 0$ , there exists  $\rho_1(\epsilon) > 1$  such that  $g_{s,i}(\eta) > \rho_1(\epsilon)\eta$  as long as  $0 < \eta \leq (1 - \epsilon)\bar{\eta}_i^*$ , and there exists  $\rho_2(\epsilon) < 1$  such that  $g_{s,i}(\eta) < \rho_2(\epsilon)\eta$  as long as  $\eta > (1 + \epsilon)\bar{\eta}_i^*$ .

An immediate consequence of Propositions 6 is the following theorem:

**Theorem 4:** Suppose  $\bar{\eta}_i^*$  is the unique positive fixed point of  $g_{s,i}(\eta)$  ( $h_{s,i}(\eta)$ , resp.), then we have

$$\bar{\eta}_i^* \geq \frac{1}{\omega_\circ^i(A, s)} \left( \frac{1}{\omega_{b\infty}^i(A^T A, s)}, \text{ resp.} \right). \quad (57)$$

Proposition 6 implies three ways to compute the fixed point  $\bar{\eta}_i^*$  for  $g_{s,i}(\eta)$ . The same discussion is also valid for  $h_{s,i}(\eta)$ .

- 1) **Naive Fixed Point Iteration:** Property 5) of Proposition 6 suggests that the fixed point iteration

$$\bar{\eta}_{t+1,i} = g_{s,i}(\bar{\eta}_{t,i}), t = 0, 1, \dots \quad (58)$$

starting from any initial point  $\bar{\eta}_{0,i}$  converges to  $\bar{\eta}_i^*$ , no matter whether  $\bar{\eta}_{0,i} < \bar{\eta}_i^*$  or  $\bar{\eta}_{0,i} > \bar{\eta}_i^*$ . Given some tolerance parameter  $\epsilon$ , the algorithm stops when  $|g_{s,i}(\bar{\eta}_{t,i}) - \bar{\eta}_{t,i}| < \epsilon$  and returns  $\bar{\eta}_i^* \approx g_{s,i}(\bar{\eta}_{t,i})$ .

The naive fixed point iteration has two major disadvantages. First, the stopping criterion based on successive improvement is not accurate as it does not reflect the gap between  $\bar{\eta}_{t,i}$  and  $\bar{\eta}_i^*$ . This disadvantage can be remedied by starting from both below and above  $\bar{\eta}_i^*$ . The distance between corresponding terms in the two generated sequences is an indication of the gap to the fixed point  $\bar{\eta}_i^*$ . However, the resulting algorithm is more computationally expensive. Second, the iteration process is slow, especially when close to the fixed point  $\eta^*$ , because  $\rho_1(\epsilon)$  and  $\rho_2(\epsilon)$  in 5) of Proposition 6 are close to 1.

- 2) **Bisection:** The bisection approach is motivated by property 4) of Proposition 6. Starting from an initial interval  $(\bar{\eta}_{L,i}, \bar{\eta}_{U,i})$  that contains  $\bar{\eta}_i^*$ , we compute  $g_{s,i}(\bar{\eta}_{M,i})$  with  $\bar{\eta}_{M,i} = (\bar{\eta}_{L,i} + \bar{\eta}_{U,i})/2$ . As a consequence of property 4),  $g_{s,i}(\bar{\eta}_{M,i}) > \bar{\eta}_{M,i}$  implies  $g_{s,i}(\bar{\eta}_{M,i}) < \bar{\eta}_i^*$ , and we set  $\bar{\eta}_{L,i} = g_{s,i}(\bar{\eta}_{M,i})$ ;  $g_{s,i}(\bar{\eta}_{M,i}) < \bar{\eta}_{M,i}$  implies  $g_{s,i}(\bar{\eta}_{M,i}) > \bar{\eta}_i^*$ , and we set  $\bar{\eta}_{U,i} = g_{s,i}(\bar{\eta}_{M,i})$ .

The convergence of the bisection approach is much faster than the naive fixed point iteration because each iteration reduces the interval length at least by half. In addition, half the length of the interval is an upper bound on the gap between  $\bar{\eta}_{M,i}$  and  $\bar{\eta}_i^*$ , resulting an accurate stopping criterion.

## VI. PRELIMINARY NUMERICAL EXPERIMENTS

In this section, we present preliminary numerical results that assess the performance of the algorithms for verifying  $\omega_2(A, s) > 0$  and computing  $\omega_2(A, s)$ . We also compare the

TABLE I  
COMPARISON OF THE SPARSITY LEVEL BOUNDS ON BLOCK-SPARSITY  
RECOVERY AND SPARSITY RECOVERY FOR A GAUSSIAN SENSING MATRIX  
 $A \in \mathbb{R}^{m \times np}$  WITH  $n = 4, p = 60$

$m$	Block Sparse Model			Sparse Model		BC	B-RIP
	$s_*$	$k_*$	$nk_*$	$s_*$	$k_*$	$k_*$	$k_*$
72	3.96	1	4	6.12	3	0	1
96	4.87	2	8	7.55	3	0	1
120	5.94	2	8	9.54	4	0	1
144	7.14	3	12	11.96	5	0	2
168	8.60	4	16	14.66	7	0	2
192	11.02	5	20	18.41	9	0	3

error bounds based on  $\omega_2(A, s)$  with the bounds based on the block RIP [2]. The involved semidefinite programs are solved using CVX.

We test the algorithms on Gaussian random matrices. The entries of Gaussian matrices are randomly generated from the standard Gaussian distribution. All  $m \times np$  matrices are normalized to have columns of unit Euclidean norm.

We first present the values of  $s_*$  computed by (34), and the values of  $k_*$  ( $k_* = \lfloor s_*/2 \rfloor$ ), and compare them with the corresponding quantities when  $A$  is used as the sensing matrix for sparsity recovery without knowing the block-sparsity structure. The quantities for the latter case are computed using the algorithms developed in [13], [16]. We also present the sparsity level bounds based on block-coherence (BC) [5] and block-RIP (B-RIP) [2]. Since we cannot compute the block-RIP constant precisely, we use Monte Carlo simulations to approximate it. We note in Table I that for the same sensing matrix  $A$ , both  $s_*$  and  $k_*$  are smaller when the block-sparse structure is exploited. However, we need to keep in mind that the true sparsity level in the block-sparse model is  $nk$ , where  $k$  is the block sparsity level. The  $nk_*$  in the fourth column for the block-sparse model is indeed much greater than the  $k_*$  in the sixth column for the sparse model, implying that exploiting the block-sparsity structure is advantageous. We also note that the computed bounds on the maximal sparsity level  $k$  are much larger than those obtained using block-coherence and block-RIP, implying that our bounds are in fact tighter for these experiments.

In the next set of experiments, we compare the computation times for the three implementation methods discussed at the end of Section V-C. The Gaussian matrix  $A$  is of size  $72 \times 120$  with  $n = 3$  and  $p = 40$ . The tolerance parameter is  $10^{-5}$ . The initial  $\eta$  value for the naive fixed point iteration is 0.1. The initial lower bound  $\eta_L$  and upper bound  $\eta_U$  are set as 0.1 and 10, respectively. All three implementations yield  $\eta^* = 0.7034$ . The CPU times for the three methods are 393 seconds, 1309 seconds, and 265 seconds. Therefore, the fixed point iteration+bisection gives the most efficient implementation in general. The bisection search method is slow in this case because the initial  $\eta_U$  is too much larger than  $\eta^*$ .

In the last experiment, we compare our recovery error bounds on the BS-BP based on  $\omega_2(A, s)$  with those based on the block

TABLE II  
 $\omega_2(A, 2k)$  AND  $\delta_{2k}(A)$  COMPUTED FOR A GAUSSIAN MATRIX  $A \in \mathbb{R}^{m \times np}$   
WITH  $n = 4$  AND  $p = 60$

$k$	$m$	72	96	120	144	168	192
		$s_*$	$k_*$	$s_*$	$k_*$	$s_*$	$k_*$
1	$\omega_2(A, 2k)$	0.45	0.53	0.57	0.62	0.65	0.67
1	$\delta_{2k}(A)$	0.90	0.79	0.66	0.58	0.55	0.51
2	$\omega_2(A, 2k)$	—	0.13	0.25	0.33	0.39	0.43
2	$\delta_{2k}(A)$	—	1.08	0.98	0.96	0.84	0.75
3	$\omega_2(A, 2k)$	—	—	—	0.11	0.18	0.25
3	$\delta_{2k}(A)$	—	—	—	1.12	1.01	0.93
4	$\omega_2(A, 2k)$	—	—	—	—	0.02	0.12
4	$\delta_{2k}(A)$	—	—	—	—	1.26	1.07
5	$\omega_2(A, 2k)$	—	—	—	—	—	0.03
5	$\delta_{2k}(A)$	—	—	—	—	—	1.28

RIP. Recall that from Corollary 2, we have for the BS-BP

$$\|\hat{x} - x\|_2 \leq \frac{2\sqrt{2k}}{\omega_2(A, 2k)} \varepsilon. \quad (59)$$

For comparison, the block RIP bounds is

$$\|\hat{x} - x\|_2 \leq \frac{4\sqrt{1 + \delta_{2k}(A)}}{1 - (1 + \sqrt{2})\delta_{2k}(A)} \varepsilon, \quad (60)$$

assuming the block RIP  $\delta_{2k}(A) < \sqrt{2} - 1$  [2]. Without loss of generality, we set  $\varepsilon = 1$ .

The block RIP is computed using Monte Carlo simulations. More explicitly, for  $\delta_{2k}(A)$ , we randomly take 1000 sub-matrices of  $A \in \mathbb{R}^{m \times np}$  of size  $m \times 2nk$  with a pattern determined by the block-sparsity structure, compute the maximal and minimal singular values  $\sigma_1$  and  $\sigma_{2k}$ , 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 block RIP is always smaller than or equal to the exact block RIP. As a consequence, the performance bounds based on the exact block RIP are *worse* than those based on the approximated block RIP. Therefore, in cases where our  $\omega_2(A, 2k)$  based bounds are better (tighter, smaller) than the approximated block RIP bounds, they are even better than the exact block RIP bounds.

In Table II, we present the values of  $\omega_2(A, 2k)$  and  $\delta_{2k}(A)$  computed for a Gaussian matrix  $A \in \mathbb{R}^{m \times np}$  with  $n = 4$  and  $p = 60$ . The corresponding  $s_*$  and  $k_*$  for different  $m$  are also included in the table. We note that in all the considered cases,  $\delta_{2k}(A) > \sqrt{2} - 1$ , and the block RIP based bound (60) does not apply at all. In contrast, the  $\omega_2$  based bound (59) is valid as long as  $k \leq k_*$ . In Table (III) we show the  $\omega_2$  based bound (59).

## VII. CONCLUSION

In this paper, we analyzed the performance of convex block-sparse signal recovery algorithms using the block- $\ell_{b\infty}$  norm of the errors as a performance criterion. We expressed other popular performance criteria in terms of the block- $\ell_{b\infty}$  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 block- $\ell_{b\infty}$  norms of the recon-

TABLE III  
THE  $\omega_2(A, 2k)$  BASED BOUNDS ON THE  $\ell_2$  NORMS OF THE ERRORS OF THE  
BS-BP FOR THE GAUSSIAN MATRIX IN TABLE II

$k$	$m$	72	96	120	144	168	192
	$s_*$	3.88	4.78	5.89	7.02	8.30	10.80
$k$	$k_*$	1	2	2	3	4	5
	$\omega_2$ bound	6.22	13.01	9.89	6.50	11.52	9.50
2	$\omega_2$ bound	—	58.56	25.37	14.64	7.30	16.26
3	$\omega_2$ bound	—	—	—	53.54	21.63	30.27
4	$\omega_2$ bound	—	—	—	—	236.74	23.25
5	$\omega_2$ bound	—	—	—	—	—	127.59

struction errors for the Block-Sparse Basis Pursuit, the Block-Sparse Dantzig Selector, and the Block-Sparse LASSO estimator. Efficient algorithms based on fixed point iteration, bisection, and semidefinite programming were implemented to 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 transmitting waveforms for compressive sensing radar.

#### APPENDIX PROOFS

##### A. Proof of Proposition 1

Many ingredients of this proof are based on well-known techniques developed for RIP and block-RIP for sparse and block-sparse recovery, respectively [2]–[4], [13]. We combine these techniques and made modifications to suit our purpose.

*Proof:* Suppose  $S = \text{bsupp}(\mathbf{x})$  and  $|S| = \|\mathbf{x}\|_{b0} = k$ . Define the error vector  $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}$ . We first prove the proposition for the BS-BP and the BS-DS. The fact that  $\|\hat{\mathbf{x}}\|_{b1} = \|\mathbf{x} + \mathbf{h}\|_{b1}$  is the minimum among all  $\mathbf{z}$ s satisfying the constraints in (5) and (6), together with the fact that the true signal  $\mathbf{x}$  satisfies the constraints as required by the conditions imposed on the noise in Proposition 1, implies that  $\|\mathbf{h}_{S^c}\|_{b1}$  cannot be very large. To see this, note that

$$\begin{aligned}
\|\mathbf{x}\|_{b1} &\geq \|\mathbf{x} + \mathbf{h}\|_{b1} \\
&= \sum_{i \in S} \|\mathbf{x}_i + \mathbf{h}_i\|_2 + \sum_{i \in S^c} \|\mathbf{x}_i + \mathbf{h}_i\|_2 \\
&\geq \sum_{i \in S} \|\mathbf{x}_i\|_2 - \sum_{i \in S} \|\mathbf{h}_i\|_2 + \sum_{i \in S^c} \|\mathbf{h}_i\|_2 \\
&= \|\mathbf{x}_S\|_{b1} - \|\mathbf{h}_S\|_{b1} + \|\mathbf{h}_{S^c}\|_{b1} \\
&= \|\mathbf{x}\|_{b1} - \|\mathbf{h}_S\|_{b1} + \|\mathbf{h}_{S^c}\|_{b1}. \tag{61}
\end{aligned}$$

Therefore, we obtain  $\|\mathbf{h}_S\|_{b1} \geq \|\mathbf{h}_{S^c}\|_{b1}$ , which leads to

$$2 \|\mathbf{h}_S\|_{b1} \geq \|\mathbf{h}_S\|_{b1} + \|\mathbf{h}_{S^c}\|_{b1} = \|\mathbf{h}\|_{b1}. \tag{62}$$

We now turn to the BS-LASSO (7). Since the noise  $\mathbf{w}$  satisfies  $\|A^T \mathbf{w}\|_{b\infty} \leq \kappa \mu$  for some  $\kappa \in (0, 1)$ , and  $\hat{\mathbf{x}}$  is the minimizer

of (7), we have

$$\frac{1}{2} \|A\hat{\mathbf{x}} - \mathbf{y}\|_2^2 + \mu \|\hat{\mathbf{x}}\|_{b1} \leq \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_2^2 + \mu \|\mathbf{x}\|_{b1}.$$

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

$$\begin{aligned}
\mu \|\hat{\mathbf{x}}\|_{b1} &\leq \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_2^2 - \frac{1}{2} \|A\hat{\mathbf{x}} - \mathbf{y}\|_2^2 + \mu \|\mathbf{x}\|_{b1} \\
&= \frac{1}{2} \|\mathbf{w}\|_2^2 - \frac{1}{2} \|A(\hat{\mathbf{x}} - \mathbf{x}) - \mathbf{w}\|_2^2 + \mu \|\mathbf{x}\|_{b1} \\
&= \frac{1}{2} \|\mathbf{w}\|_2^2 - \frac{1}{2} \|A(\hat{\mathbf{x}} - \mathbf{x})\|_2^2 \\
&\quad + \langle A(\hat{\mathbf{x}} - \mathbf{x}), \mathbf{w} \rangle - \frac{1}{2} \|\mathbf{w}\|_2^2 + \mu \|\mathbf{x}\|_{b1} \\
&\leq \langle A(\hat{\mathbf{x}} - \mathbf{x}), \mathbf{w} \rangle + \mu \|\mathbf{x}\|_{b1} \\
&= \langle \hat{\mathbf{x}} - \mathbf{x}, A^T \mathbf{w} \rangle + \mu \|\mathbf{x}\|_{b1}.
\end{aligned}$$

Using the Cauchy-Swcharz type inequality, we get

$$\begin{aligned}
\mu \|\hat{\mathbf{x}}\|_{b1} &\leq \|\hat{\mathbf{x}} - \mathbf{x}\|_{b1} \|A^T \mathbf{w}\|_{b\infty} + \mu \|\mathbf{x}\|_{b1} \\
&= \kappa \mu \|\mathbf{h}\|_{b1} + \mu \|\mathbf{x}\|_{b1},
\end{aligned}$$

which leads to

$$\|\hat{\mathbf{x}}\|_{b1} \leq \kappa \|\mathbf{h}\|_{b1} + \|\mathbf{x}\|_{b1}.$$

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

$$\begin{aligned}
&\|\mathbf{x}\|_{b1} \\
&\geq \|\hat{\mathbf{x}}\|_{b1} - \kappa \|\mathbf{h}\|_{b1} \\
&= \|\mathbf{x} + \mathbf{h}_{S^c} + \mathbf{h}_S\|_{b1} - \kappa (\|\mathbf{h}_{S^c} + \mathbf{h}_S\|_{b1}) \\
&\geq \|\mathbf{x} + \mathbf{h}_{S^c}\|_{b1} - \|\mathbf{h}_S\|_{b1} - \kappa (\|\mathbf{h}_{S^c}\|_{b1} + \|\mathbf{h}_S\|_{b1}) \\
&= \|\mathbf{x}\|_{b1} + (1 - \kappa) \|\mathbf{h}_{S^c}\|_{b1} - (1 + \kappa) \|\mathbf{h}_S\|_{b1},
\end{aligned}$$

where  $S = \text{bsupp}(\mathbf{x})$ . Consequently, we have

$$\|\mathbf{h}_S\|_{b1} \geq \frac{1 - \kappa}{1 + \kappa} \|\mathbf{h}_{S^c}\|_{b1}.$$

Therefore, similar to (62), we obtain

$$\begin{aligned}
\frac{2}{1 - \kappa} \|\mathbf{h}_S\|_{b1} &= \frac{1 + \kappa}{1 - \kappa} \|\mathbf{h}_S\|_{b1} + \frac{1 - \kappa}{1 - \kappa} \|\mathbf{h}_S\|_{b1} \\
&\geq \frac{1 + \kappa}{1 - \kappa} \frac{1 - \kappa}{1 + \kappa} \|\mathbf{h}_{S^c}\|_{b1} + \frac{1 - \kappa}{1 - \kappa} \|\mathbf{h}_S\|_{b1} \\
&= \|\mathbf{h}\|_{b1}. \tag{63}
\end{aligned}$$

■

##### B. Proof of Corollary 1

*Proof:* Suppose  $S = \text{bsupp}(\mathbf{x})$ . According to Proposition 1, we have

$$\|\mathbf{h}\|_{b1} \leq c \|\mathbf{h}_S\|_{b1} \leq ck \|\mathbf{h}\|_{b\infty}. \tag{64}$$



To prove (13), we note

$$\begin{aligned} \frac{\|\mathbf{h}\|_2^2}{\|\mathbf{h}\|_{b\infty}^2} &= \sum_{i=1}^p \left( \frac{\|\mathbf{h}_i\|_2}{\|\mathbf{h}\|_{b\infty}} \right)^2 \leq \sum_{i=1}^p \left( \frac{\|\mathbf{h}_i\|_2}{\|\mathbf{h}\|_{b\infty}} \right) \\ &= \frac{\|\mathbf{h}\|_{b1}}{\|\mathbf{h}\|_{b\infty}} \leq ck. \end{aligned} \quad (65)$$

For the first inequality, we have used  $\frac{\|\mathbf{h}_i\|_2}{\|\mathbf{h}\|_{b\infty}} \leq 1$  and  $a^2 \leq a$  for  $a \in [0, 1]$ .

For the last assertion, note that if  $\|\mathbf{h}\|_{b\infty} \leq \beta/2$ , then we have for  $i \in S$ ,

$$\|\hat{\mathbf{x}}_i\|_2 = \|\mathbf{x}_i + \mathbf{h}_i\|_2 \geq \|\mathbf{x}_i\|_2 - \|\mathbf{h}_i\|_2 > \beta - \beta/2 = \beta/2; \quad (66)$$

and for  $i \notin S$ ,

$$\|\hat{\mathbf{x}}_i\|_2 = \|\mathbf{x}_i + \mathbf{h}_i\|_2 = \|\mathbf{h}_i\|_2 < \beta/2. \quad (67)$$

### C. Proof of Theorem 1

*Proof:* Observe that for the BS-BP

$$\begin{aligned} \|A(\hat{\mathbf{x}} - \mathbf{x})\|_2 &\leq \|\mathbf{y} - A\hat{\mathbf{x}}\|_2 + \|\mathbf{y} - A\mathbf{x}\|_2 \\ &\leq \varepsilon + \|A\mathbf{w}\|_2 \leq 2\varepsilon, \end{aligned} \quad (68)$$

and similarly,

$$\|A^T A(\hat{\mathbf{x}} - \mathbf{x})\|_{b\infty} \leq 2\mu \quad (69)$$

for the BS-DS, and

$$\|A^T A(\hat{\mathbf{x}} - \mathbf{x})\|_{b\infty} \leq (1 + \kappa)\mu \quad (70)$$

for the BS-LASSO. Here for the BS-LASSO, we have used

$$\|A^T (A\hat{\mathbf{x}} - \mathbf{y})\|_{b\infty} \leq \mu, \quad (71)$$

a consequence of the optimality condition

$$A^T (A\hat{\mathbf{x}} - \mathbf{y}) \in \mu \partial \|\hat{\mathbf{x}}\|_{b1} \quad (72)$$

and the fact that the  $i$ th block of any subgradient in  $\partial \|\hat{\mathbf{x}}\|_{b1}$  is  $\hat{\mathbf{x}}_i / \|\hat{\mathbf{x}}_i\|_2$  if  $\hat{\mathbf{x}}_i \neq 0$  and is  $\mathbf{g}$  otherwise for some  $\|\mathbf{g}\|_2 \leq 1$ .

The conclusions of Theorem 1 follow from (12) and Definition 1.  $\blacksquare$

### D. Proof of Lemma 1

*Proof:* For any  $\mathbf{z} \in \{\|\mathbf{z}\|_{b\infty} = 1, \|\mathbf{z}\|_{b1} \leq s\}$ , we have

$$\mathbf{z}^T A \mathbf{z} \leq \|\mathbf{z}\|_{b1} \|A^T A \mathbf{z}\|_{b\infty} \leq s \|A^T A \mathbf{z}\|_{b\infty}. \quad (73)$$

Taking the minima of both sides of (73) over  $\{\mathbf{z} : \|\mathbf{z}\|_{b\infty} = 1, \|\mathbf{z}\|_{b1} \leq s\}$  yields

$$\omega_2^2(A, s) \leq s \omega_{b\infty}(A^T A, s). \quad (74)$$

For the other inequality, note that  $\|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_{b\infty} \leq s$  implies  $\|\mathbf{z}\|_{b1} \leq s \|\mathbf{z}\|_{b\infty} \leq s \|\mathbf{z}\|_2$ , or equivalently,

$$\{\mathbf{z} : \|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_{b\infty} \leq s\} \subseteq \{\mathbf{z} : \|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_2 \leq s\}. \quad (75)$$

As a consequence, we have

$$\begin{aligned} \omega_2(A, s) &= \min_{\|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_{b\infty} \leq s} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \frac{\|\mathbf{z}\|_2}{\|\mathbf{z}\|_{b\infty}} \\ &\geq \min_{\|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_{b\infty} \leq s} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \\ &\geq \min_{\|\mathbf{z}\|_{b1} / \|\mathbf{z}\|_2 \leq s} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \\ &= \rho_{s^2}(A), \end{aligned} \quad (76)$$

where the first inequality is due to  $\|\mathbf{z}\|_2 \geq \|\mathbf{z}\|_{b\infty}$ , and the second inequality is because the minimization is taken over a larger set.  $\blacksquare$

### E. Proof of Theorem 2

We first need some definitions. Suppose  $X$  is a scalar random variable, the Orlicz  $\psi_2$  norm of  $X$  is defined as

$$\|X\|_{\psi_2} = \inf \left\{ t > 0 : \mathbb{E} \exp \left( \frac{|X|^2}{t^2} \right) \leq 2 \right\}. \quad (77)$$

The  $\psi_2$  norm is closely related to the subgaussian constant  $L$ . We actually can equivalently define that a random vector  $\mathbf{X} \in \mathbb{R}^{np}$  is isotropic and subgaussian with constant  $L$  if  $\mathbb{E}|\langle \mathbf{X}, \mathbf{u} \rangle|^2 = \|\mathbf{u}\|_2^2$  and  $\|\langle \mathbf{X}, \mathbf{u} \rangle\|_{\psi_2} \leq L \|\mathbf{u}\|_2$  for any  $\mathbf{u} \in \mathbb{R}^{np}$  [16].

We use the notation  $\ell_*(\mathcal{H}) = \mathbb{E} \sup_{\mathbf{u} \in \mathcal{H}} \langle \mathbf{g}, \mathbf{u} \rangle$  with  $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_{np})$ , i.e., a vector of independent zero-mean unit-variance Gaussians, to denote the Gaussian width of any set  $\mathcal{H} \subset \mathbb{R}^{np}$ .

We need a result of empirical processes [16], [20]:

*Theorem 5 [16], [20]:* Let  $\{\mathbf{a}, \mathbf{a}_i, i = 1, \dots, m\} \subset \mathbb{R}^{np}$  be i.i.d. isotropic and subgaussian random vectors.  $\mathcal{H}$  is a subset of the unit sphere of  $\mathbb{R}^{np}$ , and  $\mathcal{F} = \{f_{\mathbf{u}}(\cdot) = \langle \mathbf{u}, \cdot \rangle : \mathbf{u} \in \mathcal{H}\}$ . Suppose  $\text{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) \stackrel{\text{def}}{=} \max_{f, g \in \mathcal{F}} \|f - g\|_{\psi_2} = \alpha$ . Then there exist absolute constants  $c_1, c_2, c_3$  such that for any  $\epsilon > 0$  and  $m \geq 1$  satisfying

$$m \geq c_1 \frac{\alpha^2 \ell_*^2(\mathcal{H})}{\epsilon^2}, \quad (78)$$

with probability at least  $1 - \exp(-c_2 \epsilon^2 m / \alpha^4)$ ,

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{k=1}^m f^2(\mathbf{a}_k) - \mathbb{E} f^2(\mathbf{a}) \right| \leq \epsilon. \quad (79)$$

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

$$\begin{aligned} &\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{k=1}^m f^2(\mathbf{a}_k) - \mathbb{E} f^2(\mathbf{a}) \right| \\ &\leq c_3 \max \left\{ \alpha \frac{\ell_*(\mathcal{H})}{\sqrt{m}}, \frac{\ell_*^2(\mathcal{H})}{m} \right\}. \end{aligned} \quad (80)$$

We now proceed to the proof of Theorem 2:

*Proof of Theorem 2:* We apply Theorem 5 to estimate the block  $\ell_1$ -CMSV. According to the assumptions of Theorem 2, the rows  $\{\mathbf{a}_i\}_{i=1}^m$  of  $\sqrt{m}A$  are i.i.d. isotropic and subgaussian random vectors with constant  $L$ . Consider the set  $\mathcal{H} = \{\mathbf{u} \in \mathbb{R}^{np} : \|\mathbf{u}\|_2^2 = 1, \|\mathbf{u}\|_{b1}^2 \leq s\}$  and the function set  $\mathcal{F} =$

$\{f_u(\cdot) = \langle \mathbf{u}, \cdot \rangle : \mathbf{u} \in \mathcal{H}\}$ . Clearly both  $\mathcal{H}$  and  $\mathcal{F}$  are symmetric. The diameter of  $\mathcal{F}$  satisfies

$$\alpha = \text{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) \leq 2 \sup_{\mathbf{u} \in \mathcal{H}} \|\langle \mathbf{u}, \mathbf{a} \rangle\|_{\psi_2} = 2L. \quad (81)$$

Since  $\mathbb{E} f^2(\mathbf{a}) = \mathbb{E} \langle \mathbf{u}, \mathbf{a} \rangle^2 = 1$  when  $\mathbf{a}$  follows the same distribution as  $\{\mathbf{a}_i\}_{i=1}^m$ , we observe that for any  $\epsilon \in (0, 1)$

$$\rho_s(A)^2 = \min_{\mathbf{u} \in \mathcal{H}} \mathbf{u}^T A^T A \mathbf{u} < (1 - \epsilon)^2 < (1 - \epsilon) \quad (82)$$

is a consequence of

$$\begin{aligned} & \sup_{\mathbf{u} \in \mathcal{H}} \left| \frac{1}{m} \mathbf{u}^T (\sqrt{m}A)^T (\sqrt{m}A) \mathbf{u} - 1 \right| \\ &= \sup_{\mathbf{u} \in \mathcal{H}} \left| \frac{1}{m} \sum_{i=1}^m f_u^2(\mathbf{a}_i) - \mathbb{E} f_u^2(\mathbf{a}) \right| \leq \epsilon. \end{aligned} \quad (83)$$

In view of Theorem 5, the key is to estimate the Gaussian width  $\ell_*(\mathcal{H})$

$$\begin{aligned} \ell_*(\mathcal{H}) &= \mathbb{E} \sup_{\mathbf{u}: \|\mathbf{u}\|_2=1, \|\mathbf{u}\|_{b1} \leq s} \langle \mathbf{g}, \mathbf{u} \rangle \\ &\leq \mathbb{E} \|\mathbf{u}\|_{b1} \|\mathbf{g}\|_{b\infty} \leq \sqrt{s} \mathbb{E} \|\mathbf{g}\|_{b\infty}. \end{aligned} \quad (84)$$

The quantity  $\mathbb{E} \|\mathbf{g}\|_{b\infty}$  can be bounded using Slepian's inequality ([23], Section 3.3). We rearrange the vector  $\mathbf{g} \in \mathbb{R}^{np}$  into a matrix  $G \in \mathbb{R}^{n \times p}$  such that the vectorization  $\text{vec}(G) = \mathbf{g}$ . Clearly, we have  $\|\mathbf{g}\|_{b\infty} = \|G\|_{1,2}$ , where  $\|\cdot\|_{1,2}$  denotes the matrix norm as an operator from  $(\mathbb{R}^n, \|\cdot\|_{\ell_1})$  to  $(\mathbb{R}^p, \|\cdot\|_{\ell_2})$ . Recognizing  $\|G\|_{1,2} = \max_{\mathbf{v} \in S^{n-1}, \mathbf{w} \in T^{p-1}} \langle G\mathbf{v}, \mathbf{w} \rangle$ , we define the Gaussian process  $X_{\mathbf{v},\mathbf{w}} = \langle G\mathbf{v}, \mathbf{w} \rangle$  indexed by  $(\mathbf{v}, \mathbf{w}) \in S^{n-1} \times T^{p-1}$ . Here  $S^{n-1} = \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\|_2 = 1\}$  and  $T^{p-1} = \{\mathbf{w} \in \mathbb{R}^p : \|\mathbf{w}\|_1 = 1\}$ . We compare  $X_{\mathbf{v},\mathbf{w}}$  with another Gaussian process  $Y_{\mathbf{v},\mathbf{w}} = \langle \xi, \mathbf{v} \rangle + \langle \zeta, \mathbf{w} \rangle$ ,  $(\mathbf{v}, \mathbf{w}) \in S^{n-1} \times T^{p-1}$ , where  $\xi \sim \mathcal{N}(0, I_n)$  and  $\zeta \sim \mathcal{N}(0, I_p)$ . The Gaussian processes  $X_{\mathbf{v},\mathbf{w}}$  and  $Y_{\mathbf{v},\mathbf{w}}$  satisfy the conditions for Slepian's inequality (See the proof of ([24], Theorem 32, page 23)). Therefore, we have

$$\begin{aligned} \mathbb{E} \|\mathbf{g}\|_{b\infty} &= \mathbb{E} \max_{(\mathbf{v}, \mathbf{w}) \in S^{n-1} \times T^{p-1}} X_{\mathbf{v},\mathbf{w}} \\ &\leq \mathbb{E} \max_{(\mathbf{v}, \mathbf{w}) \in S^{n-1} \times T^{p-1}} Y_{\mathbf{v},\mathbf{w}} \\ &= \mathbb{E} \max_{\mathbf{v} \in S^{n-1}} \langle \xi, \mathbf{v} \rangle + \mathbb{E} \max_{\mathbf{w} \in T^{p-1}} \langle \zeta, \mathbf{w} \rangle \\ &= \mathbb{E} \|\xi\|_2 + \mathbb{E} \|\zeta\|_\infty \\ &\leq \sqrt{n} + \sqrt{\log p}. \end{aligned} \quad (85)$$

Here we have used

$$\mathbb{E} \|\xi\|_2 \leq \sqrt{\mathbb{E} \|\xi\|_2^2} = \sqrt{n} \quad (86)$$

due to Jensen's inequality, and

$$\mathbb{E} \|\zeta\|_\infty = \mathbb{E} \max_i \zeta_i \leq \sqrt{\log p}, \quad (87)$$

a fact given by ([23], (3.13), page 79).

The conclusion of Theorem 2 then follows from (84), (85), and suitable choice of  $c_1$ . ■

## F. Proof of Proposition 2

*Proof:* We rewrite the optimization (33) as

$$\frac{1}{s^*} = \max_{\mathbf{z}} \|\mathbf{z}\|_{b\infty} \text{ s.t. } Q\mathbf{z} = 0, \|\mathbf{z}\|_{b1} \leq 1. \quad (88)$$

Note that in (88), we are maximizing a convex function over a convex set, which is in general very difficult. We will use a relaxation technique to compute an upper bound on the optimal value of (88). Define a matrix variable  $P$  of the same size as  $Q$ . Since the dual norm of  $\|\cdot\|_{b\infty}$  is  $\|\cdot\|_{b1}$ , we have

$$\begin{aligned} & \max_{\mathbf{z}} \{\|\mathbf{z}\|_{b\infty} : \|\mathbf{z}\|_{b1} \leq 1, Q\mathbf{z} = 0\} \\ &= \max_{\mathbf{u}, \mathbf{z}} \{\mathbf{u}^T \mathbf{z} : \|\mathbf{z}\|_{b1} \leq 1, \|\mathbf{u}\|_{b1} \leq 1, Q\mathbf{z} = 0\} \\ &= \max_{\mathbf{u}, \mathbf{z}} \{\mathbf{u}^T (\mathbf{z} - P^T Q \mathbf{z}) : \|\mathbf{z}\|_{b1} \leq 1, \\ & \quad \|\mathbf{u}\|_{b1} \leq 1, Q\mathbf{z} = 0\} \\ &\leq \max_{\mathbf{u}, \mathbf{z}} \{\mathbf{u}^T (I_{np} - P^T Q) \mathbf{z} : \|\mathbf{z}\|_{b1} \leq 1, \|\mathbf{u}\|_{b1} \leq 1\}. \end{aligned} \quad (89)$$

In the last expression, we have dropped the constraint  $Q\mathbf{z} = 0$ . Note that the unit ball  $\{\mathbf{z} : \|\mathbf{z}\|_{b1} \leq 1\} \subset \mathbb{R}^{np}$  is the convex hull of  $\{\mathbf{e}_i^p \otimes \mathbf{v} : 1 \leq i \leq p, \mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|_2 \leq 1\}$  and  $\mathbf{u}^T (I_{np} - P^T Q) \mathbf{z}$  is convex (actually, linear) in  $\mathbf{z}$ . As a consequence, we have

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{z}} \{\mathbf{u}^T (I_{np} - P^T Q) \mathbf{z} : \|\mathbf{z}\|_{b1} \leq 1, \|\mathbf{u}\|_{b1} \leq 1\} \\ &= \max_{j, \mathbf{v}, \mathbf{u}} \{\mathbf{u}^T (I_{np} - P^T Q) (\mathbf{e}_j^p \otimes \mathbf{v}) : \|\mathbf{u}\|_{b1} \leq 1, \\ & \quad \|\mathbf{v}\|_2 \leq 1\} \\ &= \max_j \max_{\mathbf{v}, \mathbf{u}} \{\mathbf{u}^T (I_{np} - P^T Q)_j \mathbf{v} : \|\mathbf{u}\|_{b1} \leq 1, \|\mathbf{v}\|_2 \leq 1\} \\ &= \max_j \max_{\mathbf{u}} \{\|(I_{np} - P^T Q)_j^T \mathbf{u}\|_2 : \|\mathbf{u}\|_{b1} \leq 1\}, \end{aligned} \quad (90)$$

where  $(I_{np} - P^T Q)_j$  denotes the  $j$ th column blocks of  $I_{np} - P^T Q$ , namely, the submatrix of  $I_{np} - P^T Q$  formed by the  $((j-1)n+1)$ th to  $j$ th columns.

Applying the same argument to the unit ball  $\{\mathbf{u} : \|\mathbf{u}\|_{b1} \leq 1\}$  and the convex function  $\|(I_{np} - P^T Q)_j^T \mathbf{u}\|_2$ , we obtain

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{z}} \{\mathbf{u}^T (I_{np} - P^T Q) \mathbf{z} : \|\mathbf{z}\|_{b1} \leq 1, \|\mathbf{u}\|_{b1} \leq 1\} \\ &= \max_{i,j} \|(I_{np} - P^T Q)_{i,j}\|_2. \end{aligned} \quad (91)$$

Here  $(I_{np} - P^T Q)_{i,j}$  is the submatrix of  $I_{np} - P^T Q$  formed by the  $((i-1)n+1)$ th to  $i$ th rows and the  $((j-1)n+1)$ th to  $j$ th columns, and  $\|\cdot\|_2$  is the spectral norm (the largest singular value).

Since  $P$  is arbitrary, the tightest upper bound is obtained by minimizing  $\max_{i,j} \| (I_{np} - P^T Q)_{i,j} \|_2$  with respect to  $P$ :

$$\begin{aligned} 1/s^* &= \max_z \{ \|z\|_{b\infty} : \|z\|_{b1} \leq 1, Qz = 0 \} \\ &\leq \min_P \max_{i,j} \| (I_{np} - P^T Q)_{i,j} \|_2 \\ &\stackrel{\text{def}}{=} 1/s^*. \end{aligned} \quad (92)$$

Partition  $P$  and  $Q$  as  $P = [P_1, \dots, P_p]$  and  $Q = [Q_1, \dots, Q_p]$ , with  $P_i$  and  $Q_i$  having  $n$  columns each. We explicitly write

$$(I_{np} - P^T Q)_{i,j} = \delta_{ij} I_n - P_i^T Q_j, \quad (93)$$

where  $\delta_{ij} = 1$  for  $i = j$  and 0 otherwise. As a consequence, we obtain

$$\begin{aligned} &\min_P \max_{i,j} \| (I_{np} - P^T Q)_{i,j} \|_2 \\ &= \min_{P_1, \dots, P_p} \max_i \max_j \| \delta_{ij} I_n - P_i^T Q_j \|_2 \\ &= \max_i \min_{P_i} \max_j \| \delta_{ij} I_n - P_i^T Q_j \|_2. \end{aligned} \quad (94)$$

We have moved the  $\max_i$  to the outmost because for each  $i$ ,  $\max_j \| \delta_{ij} I_n - P_i^T Q_j \|_2$  is a function of only  $P_i$  and does not depend on other variables  $P_l, l \neq i$ . ■

#### G. Proof of Proposition 4

*Proof:* We start with the continuity of  $f_{s,i}(\eta)$ . Since in the optimization problem defining  $f_{s,i}(\eta)$ , the objective function  $\|z_i\|_2$  is continuous, and the constraint correspondence

$$\begin{aligned} C(\eta) : [0, \infty) &\rightarrow \mathbb{R}^{np} \\ \eta &\mapsto \{z : \|Qz\|_{\diamond} \leq 1, \|z\|_{b1} \leq s\eta\} \end{aligned} \quad (95)$$

is compact-valued and continuous (both upper and lower hemicontinuous), according to Berge's Maximum Theorem [25] the optimal value function  $f_{s,i}(\eta)$  is continuous.

Next we show  $f_{s,i}(\eta) > s\eta$  for sufficiently small  $\eta > 0$ . Take  $z$  as the vector whose first element of the  $i$ th block is  $s\eta$  and zero otherwise. We have  $\|z\|_{b1} = s\eta$  and  $\|z_i\|_2 = s\eta > \eta$  (recall  $s \in (1, s^*)$ ). In addition, when  $\eta > 0$  is sufficiently small, we also have  $\|Qz\|_{\diamond} \leq 1$ . Therefore, for sufficiently small  $\eta$ , we have  $f_{s,i}(\eta) \geq s\eta > \eta$ .

We next prove the existence of  $\eta_B > 0$  and  $\rho_B \in (0, 1)$  such that

$$f_{s,i}(\eta) < \rho_B \eta, \quad \forall \eta > \eta_B. \quad (96)$$

We use contradiction to prove this statement. Suppose for all  $\eta_B > 0$  and  $\rho_B \in (0, 1)$ , there exists  $\eta > \eta_B$  such that  $f_{s,i}(\eta) \geq \rho_B \eta$ . Construct sequences  $\{\eta^{(k)}\}_{k=1}^{\infty} \subset (0, \infty)$ ,  $\{\rho^{(k)}\}_{k=1}^{\infty} \subset (0, 1)$ , and  $\{z^{(k)}\}_{k=1}^{\infty} \subset \mathbb{R}^{np}$  such that

$$\begin{aligned} \lim_{k \rightarrow \infty} \eta^{(k)} &= \infty, \quad \lim_{k \rightarrow \infty} \rho^{(k)} = 1 \\ \rho^{(k)} \eta^{(k)} &\leq f_{s,i}(\eta^{(k)}) = \|z_i^{(k)}\|_2, \\ \|Qz^{(k)}\|_{\diamond} &\leq 1, \quad \|z^{(k)}\|_{b1} \leq s\eta^{(k)}. \end{aligned} \quad (97)$$

Decompose  $z^{(k)} = z_1^{(k)} + z_2^{(k)}$ , where  $z_1^{(k)}$  is in the null space of  $Q$  and  $z_2^{(k)}$  in the orthogonal complement of the null space of  $Q$ . The sequence  $\{z_2^{(k)}\}_{k=1}^{\infty}$  is bounded since  $c\|z_2^{(k)}\| \leq \|Qz_2^{(k)}\|_{\diamond} \leq 1$  where  $c = \inf_{z: z \perp \text{null}(Q)} \|Qz\|_{\diamond} / \|z\| > 0$  and  $\|\cdot\|$  is any norm. Then  $\infty = \lim_{k \rightarrow \infty} \|z_i^{(k)}\|_2 \leq \lim_{k \rightarrow \infty} (\|z_{1,i}^{(k)}\|_2 + \|z_{2,i}^{(k)}\|_2)$  implies  $\{z_{1,i}^{(k)}\}_{k=1}^{\infty}$  is unbounded. For sufficiently large  $k$ , we proceed as follows:

$$\begin{aligned} s^* &> \frac{s}{\left(\frac{s}{s^*}\right)^{1/4}} \geq \frac{s\eta^{(k)}}{\rho^{(k)}\eta^{(k)}} \geq \frac{\|z^{(k)}\|_{b1}}{\|z_i^{(k)}\|_2} \\ &\geq \frac{\|z_1^{(k)}\|_{b1} - \|z_2^{(k)}\|_{b1}}{\|z_{1,i}^{(k)}\|_2 + \|z_{2,i}^{(k)}\|_2} \geq \left(\frac{s}{s^*}\right)^{1/4} \frac{\|z_1^{(k)}\|_{b1}}{\|z_{1,i}^{(k)}\|_2}, \end{aligned} \quad (98)$$

where the second and last inequalities hold only for sufficiently large  $k$  and the last inequality is due to the unboundedness of  $\{z_1^{(k)}\}$  and boundedness of  $\{z_2^{(k)}\}$ . As a consequence, we have

$$\frac{\|z_1^{(k)}\|_{b1}}{\|z_1^{(k)}\|_{b\infty}} \leq \frac{\|z_1^{(k)}\|_{b1}}{\|z_{1,i}^{(k)}\|_2} \leq \frac{s}{\sqrt{s^*}} = \sqrt{ss^*} < s^* \quad (99)$$

with  $Qz_1^{(k)} = 0$  which contradicts the definition of  $s^*$ .

To summarize the development so far, the function  $f_{s,i}(\eta) - \eta$  is continuous and for small enough  $\eta$

$$f_{s,i}(\eta) - \eta > s\eta - \eta > 0 \quad (100)$$

and for large enough  $\eta$

$$f_{s,i}(\eta) - \eta < \rho_B \eta - \eta < 0. \quad (101)$$

The intermediate value theorem then implies that there exist at least one positive  $\bar{\eta}_i^*$  such that  $f_{s,i}(\bar{\eta}_i^*) = \bar{\eta}_i^*$ .

Next we show the uniqueness of the positive fixed point  $\bar{\eta}_i^*$ , which is equal to  $\gamma_i^* \stackrel{\text{def}}{=} 1/\omega_{\diamond}^i(Q, s)$ . We first prove  $\gamma_i^* \geq \bar{\eta}_i^*$  for any fixed point  $\bar{\eta}_i^* = f_{s,i}(\bar{\eta}_i^*)$ . Suppose  $z^*$  achieves the optimal value of the optimization defining  $f_{s,i}(\bar{\eta}_i^*)$ , i.e.,

$$\bar{\eta}_i^* = f_{s,i}(\bar{\eta}_i^*) = \|z_i^*\|_2, \quad \|Qz^*\|_{\diamond} \leq 1, \quad \|z^*\|_{b1} \leq s\bar{\eta}_i^*. \quad (102)$$

Since  $\|z^*\|_{b1}/\|z_i^*\|_2 \leq s\bar{\eta}_i^*/\bar{\eta}_i^* \leq s$ , we have

$$\gamma_i^* \geq \frac{\|z_i^*\|_2}{\|Qz^*\|_{\diamond}} \geq \bar{\eta}_i^*. \quad (103)$$

If  $\bar{\eta}_i^* < \gamma_i^*$ , we define  $\bar{\eta}_0^i = (\bar{\eta}_i^* + \gamma_i^*)/2$  and

$$\begin{aligned} z^c &= \underset{z}{\operatorname{argmax}} \frac{s\|z_i\|_2}{\|z\|_{b1}} \text{ s.t. } \|Qz\|_{\diamond} \leq 1, \\ \|z_i\|_2 &\geq \bar{\eta}_0^i, \text{ and} \\ \rho &= \frac{s\|z_i^c\|_2}{\|z^c\|_{b1}}. \end{aligned} \quad (104)$$

Suppose  $z^{**}$  with  $\|Qz^{**}\|_{\diamond} = 1$  achieves the optimal value of the optimization defining  $\gamma_i^* = 1/\omega_{\diamond}^i(Q, s)$ . Clearly,  $\|z_i^{**}\|_2 = \gamma_i^* > \bar{\eta}_0^i$ , implying that  $z^{**}$  is a feasible point of the optimization defining  $z^c$  and  $\rho$ . So we have

$$\rho \geq \frac{s\|z_i^{**}\|_2}{\|z^{**}\|_{b1}} \geq 1. \quad (105)$$





For the second term  $\mathbf{u}^T P_i^T \mathbf{v}$  in the objective function of (114), the definition of  $\|\cdot\|_\diamond^*$ , the dual norm of  $\|\cdot\|_\diamond$ , leads to

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{v}} \{ \mathbf{u}^T P_i^T \mathbf{v} : \|\mathbf{u}\|_2 \leq 1, \|\mathbf{v}\|_\diamond \leq 1 \} \\ & \leq \max_{\mathbf{u}} \{ \|\mathbf{u}\|_\diamond^* : \|\mathbf{u}\|_2 \leq 1 \} = \|P_i\|_{2,*}, \end{aligned} \quad (116)$$

where  $\|P_i\|_{2,*}$  is the operator norm of  $P_i$  from the space with norm  $\|\cdot\|_2$  to the space with norm  $\|\cdot\|_\diamond^*$ . A tight upper bound on  $\max_{\mathbf{z}} \{ \|\mathbf{z}\|_2 : \|\mathbf{z}\|_{b1} \leq s\eta, \|Q\mathbf{z}\|_\diamond \leq 1 \}$  is obtained by minimizing with respect to  $P_i$ :

$$\begin{aligned} & \max_{\mathbf{z}} \{ \|\mathbf{z}\|_2 : \|\mathbf{z}\|_{b1} \leq s\eta, \|Q\mathbf{z}\|_\diamond \leq 1 \} \\ & \leq \min_{P_i} \max_j s\eta \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 + \|P_i\|_{2,*}. \end{aligned} \quad (117)$$

When  $Q = A$  and  $\diamond = 2$ , the operator norm

$$\|P_i\|_{2,*} = \|P_i\|_{2,2} = \|P_i\|_2. \quad (118)$$

When  $Q = A^T A$  and  $\diamond = b\infty$ , we have

$$\begin{aligned} & \|P_i\|_{2,*} = \|P_i\|_{2,b1} = \max_{\mathbf{v}: \|\mathbf{v}\|_2 \leq 1} \|P_i \mathbf{v}\|_{b1} \\ & = \max_{\mathbf{v}: \|\mathbf{v}\|_2 \leq 1} \sum_{l=1}^p \|P_i^l \mathbf{v}\|_2 \leq \sum_{l=1}^p \|P_i^l\|_2. \end{aligned} \quad (119)$$

Here  $P_i^l$  is the submatrix of  $P_i$  formed by the  $((l-1)n+1)$ th to  $ln$ th rows. Proposition 5 then follows from (117), (118), and (119). ■

### I. Proof of Proposition 6

*Proof:* We prove the proposition for  $g_{s,i}(\eta)$ . The proof for  $h_{s,i}(\eta)$  is very similar.

- 1) First note that adding an additional constraint  $\|P_i\|_2 \leq s\eta$  in the definition of  $g_{s,i}$  does not change the definition, because  $g_{s,i}(\eta) \leq s\eta$  as easily seen by setting  $P_i = 0$ :

$$\begin{aligned} g_{s,i}(\eta) &= \min_{P_i} \left\{ s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2 : \right. \\ & \quad \left. \|P_i\|_2 \leq s\eta \right\} \\ &= - \max_{P_i} \left\{ -s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) \right. \\ & \quad \left. - \|P_i\|_2 : \|P_i\|_2 \leq s\eta \right\}. \end{aligned} \quad (120)$$

Since the objective function to be maximized is continuous, and the constraint correspondence

$$C(\eta) = \{P_i : \|P_i\|_2 \leq s\eta\} \quad (121)$$

is compact-valued and continuous (both upper and lower hemicontinuous), according to Berge's Maximum Theorem [25] the optimal value function  $g_{s,i}(\eta)$  is continuous.

To show the strict increasing property, suppose  $\eta_1 < \eta_2$  and  $P_{i,2}^*$  achieves  $g_{s,i}(\eta_2)$ . Then we have

$$\begin{aligned} g_{s,i}(\eta_1) &\leq s\eta_1 \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_{i,2}^{*T} Q_j\|_2 \right) + \|P_{i,2}^*\|_2 \\ &< s\eta_2 \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_{i,2}^{*T} Q_j\|_2 \right) + \|P_{i,2}^*\|_2 \\ &= g_{s,i}(\eta_2). \end{aligned} \quad (122)$$

The concavity of  $g_{s,i}(\eta)$  follows from the fact that  $g_{s,i}(\eta)$  is the minimization of a function of variables  $\eta$  and  $P_i$ , and when  $P_i$ , the variable to be minimized, is fixed, the function is linear in  $\eta$ .

- 2) Next we show that when  $\eta > 0$  is sufficiently small  $g_{s,i}(\eta) \geq s\eta$ . For any  $i$ , we have the following,

$$\begin{aligned} g_{s,i}(\eta) &= \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2 \\ &\geq \min_{P_i} s\eta (1 - \|P_i^T Q_i\|_2) + \|P_i\|_2 \\ &\geq \min_{P_i} s\eta (1 - \|P_i\|_2 \|Q_i\|_2) + \|P_i\|_2 \\ &= s\eta + \min_{P_i} \|P_i\|_2 (1 - s\eta \|Q_i\|_2) \\ &\geq s\eta > \eta, \end{aligned} \quad (123)$$

where the minimum of the last optimization is achieved at  $P_i = 0$  when  $\eta < 1/(s \|Q_i\|_2)$ .

Recall that

$$1/s_* = \max_i \min_{P_i} \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2. \quad (124)$$

Suppose  $P_i^*$  is the optimal solution for each  $\min_{P_i} \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2$ . For every  $i$ , we then have

$$1/s_* \geq \max_j \|\delta_{ij} \mathbf{I}_n - P_i^{*T} Q_j\|_2, \quad (125)$$

which implies

$$\begin{aligned} g_{s,i}(\eta) &= \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2 \\ &\leq s\eta \left( \max_j \|\delta_{ij} \mathbf{I}_n - P_i^{*T} Q_j\|_2 \right) + \|P_i^*\|_2 \\ &\leq \frac{s}{s_*} \eta + \|P_i^*\|_2. \end{aligned} \quad (126)$$

Pick  $\rho \in (s/s_*, 1)$ . Then, we have the following when  $\eta > \|P_i^*\|_2/(\rho - s/s_*)$ :

$$g_{s,i}(\eta) \leq \rho\eta. \quad (127)$$

- 3) Properties 1) and 2) imply that  $g_{s,i}(\eta)$  has at least one positive fixed point. To prove uniqueness, suppose there are two fixed points  $0 < \bar{\eta}_{1,i} < \bar{\eta}_{2,i}$ . Pick  $\eta_0$  small enough such that  $g_{s,i}(\eta_0) > \eta_0 > 0$  and  $\eta_0 < \bar{\eta}_{1,i}$ . Then  $\bar{\eta}_{1,i} = \lambda\eta_0 + (1-\lambda)\bar{\eta}_{2,i}$  for some  $\lambda \in (0, 1)$ , which implies that  $g_{s,i}(\bar{\eta}_{1,i}) \geq \lambda g_{s,i}(\eta_0) + (1-\lambda)g_{s,i}(\bar{\eta}_{2,i}) > \lambda\eta_0 + (1-\lambda)\bar{\eta}_{2,i} = \bar{\eta}_{1,i}$  due to the concavity, contradicting  $\bar{\eta}_{1,i} = g_{s,i}(\bar{\eta}_{1,i})$ .
- 4) This property follows from the continuity, the uniqueness, and property 2).

- 5) We use contradiction to show the existence of  $\rho_1(\epsilon)$  in 5). In view of 2), we need only to show the existence of such a  $\rho_1(\epsilon)$  that works for  $\eta_L \leq \eta \leq (1 - \epsilon)\bar{\eta}_i^*$  where  $\eta_L = \sup\{\eta : g_{s,i}(\xi) > s\xi, \forall 0 < \xi \leq \eta\}$ . Supposing otherwise, we then construct sequences  $\{\eta^{(k)}\}_{k=1}^\infty \subset [\eta_L, (1 - \epsilon)\bar{\eta}_i^*]$  and  $\{\rho_1^{(k)}\}_{k=1}^\infty \subset (1, \infty)$  with

$$\lim_{k \rightarrow \infty} \rho_1^{(k)} = 1 \text{ and } g_{s,i}(\eta^{(k)}) \leq \rho^{(k)} \eta^{(k)}. \quad (128)$$

Due to the compactness of  $[\eta_L, (1 - \epsilon)\bar{\eta}_i^*]$ , there must exist a subsequence  $\{\eta^{(k_l)}\}_{l=1}^\infty$  of  $\{\eta^{(k)}\}$  such that  $\lim_{l \rightarrow \infty} \eta^{(k_l)} = \eta_{\text{lim}}$  for some  $\eta_{\text{lim}} \in [\eta_L, (1 - \epsilon)\bar{\eta}_i^*]$ . As a consequence of the continuity of  $g_{s,i}(\eta)$ , we have

$$g_{s,i}(\eta_{\text{lim}}) = \lim_{l \rightarrow \infty} g_{s,i}(\eta^{(k_l)}) \leq \lim_{l \rightarrow \infty} \rho_1^{(k_l)} \eta^{(k_l)} = \eta_{\text{lim}}. \quad (129)$$

Again due to the continuity of  $g_{s,i}(\eta)$  and that  $g_{s,i}(\eta) < \eta$  for  $\eta < \eta_L$ , there exists  $\eta_c \in [\eta_L, \eta_{\text{lim}}]$  such that

$$g_{s,i}(\eta_c) = \eta_c, \quad (130)$$

contradicting the uniqueness of the fixed point of  $g_{s,i}(\eta)$ . The existence of  $\rho_2(\epsilon)$  can be proved similarly. ■

## 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] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [3] 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.
- [4] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," in *Proc. Compte Rendus de l'Académie des Sciences, Paris, Serie I*, 2008, vol. 346, pp. 589–592.
- [5] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [6] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via  $\ell_1$  minimization," *Math. Program.*, vol. 127, no. 1, pp. 57–88, 2011.
- [7] 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.
- [8] 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.
- [9] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [10] 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.
- [11] S. Sen and A. Nehorai, "Sparsity-based multi-target tracking using OFDM radar," *IEEE Trans. Signal Process.*, vol. 59, pp. 1902–1906, Apr. 2011.
- [12] G. Tang and A. Nehorai, "Computable performance analysis of block-sparsity recovery," in *Proc. 2011 4th IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2011, pp. 265–268.
- [13] G. Tang and A. Nehorai, "Computable performance bounds on sparse recovery," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 132–141, Jan. 2015.
- [14] A. d'Aspremont, L. E. 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.
- [15] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Math. Program.*, vol. 127, no. 1, pp. 123–144, 2011.
- [16] 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.
- [17] H. Liu, J. Zhang, X. Jiang, and J. Liu, "The group dantzig selector," presented at the *Int. Conf. Artif. Intell. Statist. (AISTATS 2010)*, Sardinia, May 2010.
- [18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc.: Series B (Statist. Methodol.)*, vol. 68, no. 1, pp. 49–67, 2006.
- [19] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
- [20] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geometr. Funct. Anal.*, pp. 1248–1282, Nov. 2007.
- [21] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming (Web Page and Software)," Jun. 2009. [Online]. Available: <http://stanford.edu/~boyd/cvx>
- [22] Y. Nesterov, "Smoothing technique and its applications in semidefinite optimization," *Math. Program.*, vol. 110, pp. 245–259, Mar. 2007.
- [23] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*, Berlin, Germany: Springer-Verlag, 1991.
- [24] R. Vershynin, Y. Eldar, and G. Kutyniok, Eds., "Introduction to the non-asymptotic analysis of random matrices," *Compressed Sensing: Theory Appl.* to appear.
- [25] C. Berge, *Topological Spaces*, Mineola, NY, USA: Dover Publications, reprinted 1997 in paperback.



optimization, machine learning, and their applications in data analysis, optics, imaging, and networks.



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. Earlier, he was a faculty member at Yale University, New Haven, CT, USA, and the University of Illinois at Chicago, USA.

Dr. Nehorai served as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002. From 2003 to 2005 he was the Vice President (Publications) of the IEEE Signal Processing Society (SPS), the 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. 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 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. He 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 has been a Fellow of the Royal Statistical Society since 1996 and a Fellow of AAAS since 2012.

**Gongguo Tang** (S'09–M'11) is an Assistant Professor in the Electrical Engineering and Computer Science Department at Colorado School of Mines since 2014. He received his Ph.D. degree in electrical engineering from Washington University in St. Louis in 2011. He was a Post-doctoral Research Associate at the Department of Electrical and Computer Engineering, University of Wisconsin-Madison in 2011–2013, and a visiting scholar at University of California, Berkeley in 2013. Gongguo's research interests are in the area of signal processing, convex optimization, machine learning, and their applications in data analysis, optics, imaging, and networks.

**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, USA.

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