# Fixed point theory and semidefinite programming for computable performance analysis of block-sparsity recovery

**Gongguo Tang · Arye Nehorai**

**Abstract** In this paper, we employ 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, a computable performance bound 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 goodness 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 goodness measures. We demonstrate that as long as the number of measurements is relatively large, these goodness 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 goodness measures with the fixed points of functions defined by a series of semidefinite programs. This relation with fixed point theory yields efficient algorithms with global convergence guarantees to compute the goodness measures.

Gongguo Tang
Washington University in St. Louis, St. Louis, MO 63130-1127
E-mail: gt2@ese.wustl.edu

Arye Nehorai
Washington University in St. Louis, St. Louis, MO 63130-1127
E-mail: nehorai@ese.wustl.edu

## 1 Introduction

Recovery of signals with low-dimensional structures, in particular, sparsity [4], block-sparsity [7], and low-rankness [3], has found numerous applications in signal sampling, control, 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:

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

where $\boldsymbol{x} \in \mathbb{R}^N$ is the signal to be reconstructed, $\boldsymbol{y} \in \mathbb{R}^m$ 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. For example, a sparse signal is assumed to have only a few non-zero coefficients when represented as a linear combination of atoms from an orthogonal basis or from an overcomplete dictionary. Similarly, for a block-sparse signal, the non-zero coefficients are assumed to cluster into blocks.

A theoretically justified way to exploit the low-dimensional structure in recovering $\boldsymbol{x}$ 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 enforcements is usually analyzed using variants of the restricted isometry property (RIP) [2, 3, 7]. Upper bounds on the $\ell_2$ norm of the error vectors for various recovery algorithms have been expressed in terms of the RIP. 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 itself. Actually, the only known sensing matrices with nice RIPs are certain types of random matrices [9].

In this paper, we investigate the recovery performance for block-sparse signals. Block-sparsity arises naturally in applications such as sensor arrays [12], radar [18], multi-band signals [14], and DNA microarrays [16]. 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 [17, 18].

The aforementioned applications require a computable performance analysis of block-sparsity recovery. While it is perfectly reasonable to use a random sensing matrix for signal sampling, the sensing matrices in other applications are far from random and actually 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 goodness measures of the sensing matrices. Computable performance measures 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 analysis would form a bisis for optimal sensing matrix design.

We preview our contributions. First of all, we define a family of goodness measures of the sensing matrix, and use them to derive 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. Our preliminary numerical results show that these bounds are tighter than the block RIP based bounds. Second and most important, we develop a fixed point iteration framework to design algorithms that efficiently compute lower bounds on the goodness 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 optima from any initial point. As a by-product, we obtain a fast algorithm to verify the sufficient condition guaranteeing exact block-sparsity recovery via block-$\ell_1$ minimization. Finally, we show that the goodness 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.

This work extends verifiable and computable performance analysis from sparse signal cases [5, 6, 9, 21] 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 goodness 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 2, we introduce notations and we present the measurement model, three convex relaxation algorithms, and the sufficient and necessary condition for exact block-$\ell_1$ recovery. In section 3, we derive performance bounds on the block-$\ell_\infty$ norms of the recovery errors for several convex relaxation algorithms. Section 4 is devoted to the probabilistic analysis of our block-$\ell_\infty$ performance measures. In Section 5, we design algorithms to verify a sufficient condition for exact block-$\ell_1$ recovery in the noise-free case, and to compute the goodness measures of arbitrary sensing matrices through fixed point iteration, bisection search, and semidefinite programming. We evaluate the algorithms' performance in Section 6. Section 7 summarizes our conclusions.

## 2 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 $\boldsymbol{x} \in \mathbb{R}^{np}$, we partition the vector into $p$ blocks, each of length $n$. More precisely, we have $\boldsymbol{x} = \left[\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_p^T\right]^T$, with the $i$th block $\boldsymbol{x}_i \in \mathbb{R}^n$. The block-$\ell_q$ norms for $1 \leq q \leq \infty$ associated with this block-sparse structure are defined as

$$\|\boldsymbol{x}\|_{\mathrm{b}q} = \left(\sum_{i=1}^{p} \|\boldsymbol{x}_i\|_2^q\right)^{1/q}, \ 1 \leq q < \infty \tag{2}$$

and

$$\|\boldsymbol{x}\|_{\mathrm{b}\infty} = \max_{1 \leq i \leq p} \|\boldsymbol{x}_i\|_2, \ q = \infty. \tag{3}$$

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

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

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

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

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{w}, \tag{4}$$

where $A \in \mathbb{R}^{m \times np}$ is the measurement/sensing matrix, $\boldsymbol{y}$ is the measurement vector, and $\boldsymbol{w}$ is noise. A very special block-sparse model is when $\boldsymbol{x}$ is a complex signal, such as the models in sensor arrays and radar applications. Note that the computable performance analysis developed in [20, 21] for the real variables can not apply to the complex case.

Many algorithms in sparse signal recovery have been extended to recover the block-sparse signal $\boldsymbol{x}$ from $\boldsymbol{y}$ by exploiting the block-sparsity of $\boldsymbol{x}$. We focus on three algorithms based on block-$\ell_1$ minimization: the Block-Sparse Basis Pursuit (BS-BP) [7], the Block-Sparse Dantzig selector (BS-DS) [11],

and the Block-Sparse LASSO estimator (BS-LASSO) [24].

$$\text{BS-BP: } \min_{\boldsymbol{z}\in\mathbb{R}^{np}} \|\boldsymbol{z}\|_{\mathrm{b}1} \quad \text{s.t. } \|\boldsymbol{y} - A\boldsymbol{z}\|_2 \leq \varepsilon \tag{5}$$

$$\text{BS-DS: } \min_{\boldsymbol{z}\in\mathbb{R}^{np}} \|\boldsymbol{z}\|_{\mathrm{b}1} \quad \text{s.t. } \|A^T(\boldsymbol{y} - A\boldsymbol{z})\|_{\mathrm{b}\infty} \leq \mu \tag{6}$$

$$\text{BS-LASSO: } \min_{\boldsymbol{z}\in\mathbb{R}^{np}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 + \mu\|\boldsymbol{z}\|_{\mathrm{b}1}. \tag{7}$$

Here $\mu$ is a tuning parameter, and $\varepsilon$ is a measure of the noise level. All three optimization problems have efficient implementations using convex programming.

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

$$\min_{\boldsymbol{z}\in\mathbb{R}^{np}} \|\boldsymbol{z}\|_{\mathrm{b}1} \quad \text{s.t. } A\boldsymbol{z} = A\boldsymbol{x}, \tag{8}$$

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

$$\min_{\boldsymbol{z}\in\mathbb{R}^{n}} \|\boldsymbol{z}\|_{\mathrm{b}0} \quad \text{s.t. } A\boldsymbol{z} = A\boldsymbol{x}. \tag{9}$$

A minimal requirement on the block-$\ell_1$ minimization algorithms is the *uniqueness and exactness* of the solution $\hat{\boldsymbol{x}} \stackrel{\text{def}}{=} \operatorname{argmin}_{\boldsymbol{z}:A\boldsymbol{z}=A\boldsymbol{x}}\|\boldsymbol{x}\|_{\mathrm{b}1}$, *i.e.*, $\hat{\boldsymbol{x}} = \boldsymbol{x}$. When the true signal $\boldsymbol{x}$ is $k-$block-sparse, the sufficient and necessary condition for exact block-$\ell_1$ recovery is [19]

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

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

## 3 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 8.1.

**Proposition 1** *Suppose $\boldsymbol{x} \in \mathbb{R}^{np}$ in (4) is $k-$block-sparse and the noise $\boldsymbol{w}$ satisfies $\|\boldsymbol{w}\|_2 \leq \varepsilon$, $\|A^T\boldsymbol{w}\|_{\mathrm{b}\infty} \leq \mu$, and $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\mu, \kappa \in (0,1)$, for the BS-BP, the BS-DS, and the BS-LASSO, respectively. Define $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ as the error vector for any of the three block-$\ell_1$ recovery algorithms (5), (6), and (7). Then we have*

$$\|\boldsymbol{h}_S\|_{\mathrm{b}1} \geq \|\boldsymbol{h}\|_{\mathrm{b}1}/c, \tag{11}$$

*where $S = \operatorname{bsupp}(\boldsymbol{x})$, $c = 2$ 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 8.2.

**Corollary 1** *Under the assumptiocorollaryns of Proposition 1, we have*

$$\|\boldsymbol{h}\|_{\mathrm{b}1} \leq ck\|\boldsymbol{h}\|_{\mathrm{b}\infty}, \ and \tag{12}$$

$$\|\boldsymbol{h}\|_2 \leq \sqrt{ck}\|\boldsymbol{h}\|_{\mathrm{b}\infty}. \tag{13}$$

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

$$\{i : \|\hat{\boldsymbol{x}}_i\|_2 > \beta/2\} = \mathrm{bsupp}(\boldsymbol{x}), \tag{14}$$

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

For ease of presentation, we introduce the following notation:

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

$$\omega_\diamond(Q,s) = \min_{\boldsymbol{z}:\|\boldsymbol{z}\|_{\mathrm{b}1}/\|\boldsymbol{z}\|_{\mathrm{b}\infty} \leq s} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_{\mathrm{b}\infty}}, \tag{15}$$

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 8.3.

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

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

*for the BS-BP,*

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

*for the BS-DS, and*

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

*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*

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

*for the BS-BP,*

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

*for the BS-DS, and*

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

*for the BS-LASSO.*

One of the primary contributions of this work is the design of algorithms that compute $\omega_\diamond(A, s)$ and $\omega_{\mathrm{b}\infty}(A^T A, s)$ efficiently. 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 4, we also 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 6 show that in many cases the error bounds on the $\ell_2$ norms based on Corollary 2 are tighter than the block RIP based bounds. Before we turn to the computation issues, we first establish some results on the probability behavior of $\omega_\diamond(Q, s)$.

## 4 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 [20]:

**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

$$\rho_s(A) = \min_{\boldsymbol{z}:\ \|\boldsymbol{z}\|_{\mathrm{b}1}^2/\|\boldsymbol{z}\|_2^2 \leq s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}. \tag{22}$$

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 $\omega_{b\infty}(A^TA,s)$, 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.

**Lemma 1**

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

For a proof, see Appendix 8.4.

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

**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*

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

*we have*

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

*and*

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

For a proof, see Appendix 8.5.

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

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2\varepsilon}{\rho_{2k}(A)}. \tag{27}$$

The conclusion of Theorem 2, combined with the previous equation, implies that we could stably recover a block sparse signal 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 $kn$ [20]. 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 [7, 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*

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

*we have*

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

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

*and*

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

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

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 5. 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 $2 \leq q \leq \infty$ [13].

## 5 Verification and Computation of $\omega_\diamond$

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

### 5.1 Verification of $\omega_\diamond > 0$

A prerequisite for the bounds in Theorem 1 to be valid is the positiveness of the involved $\omega_\diamond(\cdot)$. 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 block-$\ell_1$ recovery.

Verifying $\omega_\diamond(Q, s) > 0$ amounts to making sure $\|\boldsymbol{z}\|_{\mathrm{b}1}/\|\boldsymbol{z}\|_{\mathrm{b}\infty} \leq s$ for all $\boldsymbol{z}$ such that $Q\boldsymbol{z} = 0$. Therefore, we compute

$$s^* = \min_{\boldsymbol{z}} \frac{\|\boldsymbol{z}\|_{\mathrm{b}1}}{\|\boldsymbol{z}\|_{\mathrm{b}\infty}} \text{ s.t. } Q\boldsymbol{z} = 0. \tag{33}$$

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** *The reciprocal, denoted by $s_*$, of the optimal value of the following optimization,*

$$\max_i \min_{P_i} \max_j \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2 \tag{34}$$

*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 $P = [P_1, \ldots, P_p]$, $Q = [Q_1, \ldots, Q_p]$ with $P_i$ and $Q_j$ having $n$ columns each.*

The proof is given in Appendix 8.6.

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

$$\min_{P_i} \max_j \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2. \tag{35}$$

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

$$\min_{P_i, t} t \text{ s.t. } \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2 \leq t, j = 1, \ldots, p, \tag{36}$$

$$\Leftrightarrow \min_{P_i, t} t \text{ s.t. } \begin{bmatrix} t\mathrm{I}_n & \delta_{ij}\mathrm{I}_n - P_i^T Q_j \\ \delta_{ij}I_n - Q_j^T P_i & t\mathrm{I}_n \end{bmatrix} \succeq 0, j = 1, \ldots, p. \tag{37}$$

Small instances of (36) and (37) can be solved using CVX [8]. However, for medium to large scale problems, it is beneficial to use first-order techniques to solve (35) directly. We observe that $\max_j \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2$ can be expressed as the largest eigenvalue of a block-diagonal matrix. The smoothing technique for semidefinite optimization developed in [15] can be used to minimize the largest eigenvalue with respect to $P_i$. We leave these implementations 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 $A^T$.

5.2 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:

$$\omega_\diamond(Q, s) = \min_{\boldsymbol{z}} \frac{\|Q\boldsymbol{z}\|_\diamond}{\|\boldsymbol{z}\|_{\mathrm{b}\infty}} \text{ s.t. } \frac{\|\boldsymbol{z}\|_{\mathrm{b}1}}{\|\boldsymbol{z}\|_{\mathrm{b}\infty}} \leq s, \tag{38}$$

or equivalently,

$$\frac{1}{\omega_\diamond(Q,s)} = \max_{z} \|z\|_{b\infty} \text{ s.t. } \|Qz\|_\diamond \leq 1, \frac{\|z\|_{b1}}{\|z\|_{b\infty}} \leq s. \tag{39}$$

For any $s \in (1, s^*)$, we define a function over $[0, \infty)$ parameterized by $s$

$$f_s(\eta) = \max_{z} \{\|z\|_{b\infty} : \|Qz\|_\diamond \leq 1, \|z\|_{b1} \leq s\eta\}. \tag{40}$$

We basically replaced the $\|z\|_{b\infty}$ in the denominator of the fractional constraint in (39) with $\eta$. It turns out that the unique positive fixed point of $f_s(\eta)$ is exactly $1/\omega_\diamond(Q,s)$, as shown by the following proposition. See Appendix 8.7 for the proof.

**Proposition 3** *The function $f_s(\eta)$ has the following properties:*

1. *$f_s(\eta)$ is continuous in $\eta$.*
2. *$f_s(\eta)$ is strictly increasing in $\eta$.*
3. *$f_s(0) = 0$, $f_s(\eta) \geq s\eta > \eta$ for sufficiently small $\eta > 0$, and there exists $\rho < 1$ such that $f_s(\eta) < \rho\eta$ for sufficiently large $\eta$.*
4. *$f_s(\eta)$ has a unique positive fixed point $\eta^* = f_s(\eta^*)$ that is equal to $1/\omega_\diamond(Q,s)$.*
5. *For $\eta \in (0, \eta^*)$, we have $f_s(\eta) > \eta$, and for $\eta \in (\eta^*, \infty)$, we have $f_s(\eta) < \eta$.*
6. *For any $\epsilon > 0$, there exists $\rho_1(\epsilon) > 1$ such that $f_s(\eta) > \rho_1(\epsilon)\eta$ as long as $0 < \eta < (1-\epsilon)\eta^*$, and there exists $\rho_2(\epsilon) < 1$ such that $f_s(\eta) < \rho_2(\epsilon)\eta$ as long as $\eta > (1+\epsilon)\eta^*$.*

We have transformed the problem of computing $\omega_\diamond(Q,s)$ into one of finding the positive fixed point of a one-dimensional function $f_s(\eta)$. Property 6) of Proposition 3 states that we could start with any $\eta_0$ and use the iteration

$$\eta_{t+1} = f_s(\eta_t), t = 0, 1, \cdots \tag{41}$$

to find the positive fixed point $\eta^*$. In addition, if we start from two initial points, one less than $\eta^*$ and one greater than $\eta^*$, then the gap between the generated sequences indicates how close we are to the fixed point $\eta^*$.

Property 5) suggests finding $\eta^*$ by bisection search. Suppose we have an interval $(\eta_L, \eta_U)$ that includes $\eta^*$. Consider the middle point $\eta_M = \frac{\eta_L + \eta_U}{2}$. If $f_s(\eta_M) < \eta_M$, we conclude that $\eta^* < \eta_M$ and we set $\eta_U = f(\eta_M)$; if $f_s(\eta_M) > \eta_M$, we conclude that $\eta^* > \eta_M$ and we set $\eta_L = f(\eta_M)$. We continue this bisection procedure until the interval length $\eta_U - \eta_L$ is sufficiently small.

5.3 Relaxation of the Subproblem

Unfortunately, except when $n = 1$ and the signal is real, *i.e.,* the real sparse case, it is not easy to compute $f_s(\eta)$ according to (40). In the following theorem, we present a relaxation of the subproblem

$$\max_{z} \|z\|_{b\infty} \text{ s.t. } \|Qz\|_\diamond \leq 1, \|z\|_{b1} \leq s\eta \tag{42}$$

by computing an upper bound on $f_s(\eta)$. This proof is similar to that of Proposition 2 and is given in Appendix 8.8.

**Proposition 4** *When $Q = A$ and $\diamond = 2$, we have*

$$f_s(\eta) \leq \max_i \min_{P_i} \max_j s\eta \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 + \|P_i\|_2, and \tag{43}$$

*when $Q = A^T A$ and $\diamond = \mathrm{b}\infty$, we have*

$$f_s(\eta) \leq \max_i \min_{P_i} \max_j s\eta \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 + \sum_{l=1}^{p} \|P_i^l\|_2. \tag{44}$$

*Here $P_i$ (resp. $Q_j$) is the submatrix of $P$ (resp. $Q$) formed by the $(i-1)n+1th$ to $inth$ columns (resp. $(j-1)n+1th$ to $jnth$ columns), and $P_i^l$ is the submatrix of $P$ formed by the $(i-1)n+1th$ to $inth$ columns and the $(l-1)n+1th$ to $lnth$ rows.*

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

$$\min_{P_i} \max_j s\eta \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 + \|P_i\|_2 \tag{45}$$

can be solved using semidefinite programming:

$$\min_{P_i, t_0, t_1} s\eta t_0 + t_1 \text{ s.t. } \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 \leq t_0, j = 1, \ldots, p, \|P_i\|_2 \leq t_1,$$

$$\Leftrightarrow \min_{P_i, t_0, t_1} s\eta t_0 + t_1 \text{ s.t.}$$

$$\begin{bmatrix} t_0 \mathrm{I}_n & \delta_{ij} \mathrm{I}_n - P_i^T Q_j \\ \delta_{ij} \mathrm{I}_n - Q_j^T P_i & t_0 \mathrm{I}_n \end{bmatrix} \succeq 0, j = 1, \ldots, p,$$

$$\begin{bmatrix} t_1 \mathrm{I}_m & P_i \\ P_i^T & t_1 \mathrm{I}_n \end{bmatrix} \succeq 0. \tag{46}$$

Similarly, the optimization problem

$$\min_{P_i} \max_j s\eta \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 + \sum_{l=1}^{p} \|P_i^l\|_2 \tag{47}$$

can be solved by the following semidefinite program:

$$\min_{P_i, t_0, t_1, \ldots, t_p} s\eta t_0 + \sum_{l=1}^{p} t_l \text{ s.t. } \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 \leq t_0, j = 1, \ldots, p, \|P_i^l\|_2 \leq t_l, l = 1, \ldots, p,$$

$$\Leftrightarrow \min_{P_i, t_0, t_1} s\eta t_0 + \sum_{l=1}^{p} t_l \text{ s.t.}$$

$$\begin{bmatrix} t_0 \mathrm{I}_n & \delta_{ij} \mathrm{I}_n - P_i^T Q_j \\ \delta_{ij} \mathrm{I}_n - Q_j^T P_i & t_0 \mathrm{I}_n \end{bmatrix} \succeq 0, j = 1, \ldots, p,$$

$$\begin{bmatrix} t_l \mathrm{I}_n & P_i^l \\ P_i^{lT} & t_l \mathrm{I}_n \end{bmatrix} \succeq 0, l = 1, \ldots, p. \tag{48}$$

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

5.4 Fixed Point Theory for Computing a Lower Bound on $\omega_\diamond$

Although Proposition 4 provides ways to efficiently compute upper bounds on the subproblem (42) for fixed $\eta$, it is not obvious whether we could use it to compute an upper bound on the positive fixed point of $f_s(\eta)$, or $1/\omega_\diamond(Q, s)$. We show in this subsection that another iterative procedure can compute such upper bounds.

To this end, we define functions $g_{s,i}(\eta)$ and $g_s(\eta)$ over $[0, \infty)$, parameterized by $s$ for $s \in (1, s_*)$,

$$g_{s,i}(\eta) = \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2, \text{ and}$$
$$g_s(\eta) = \max_i g_{s,i}(\eta). \tag{49}$$

The following proposition, whose proof is given in Appendix 8.9, lists some properties of $g_{s,i}(\eta)$ and $g_s(\eta)$.

**Proposition 5** *The functions $g_{s,i}(\eta)$ and $g_s(\eta)$ have the following properties:*

1. *$g_{s,i}(\eta)$ and $g_s(\eta)$ are continuous in $\eta$.*
2. *$g_{s,i}(\eta)$ and $g_s(\eta)$ are strictly increasing in $\eta$.*
3. *$g_{s,i}(\eta)$ is concave for every $i$.*
4. *$g_s(0) = 0$, $g_s(\eta) \geq s\eta > \eta$ for sufficiently small $\eta > 0$, and there exists $\rho < 1$ such that $g_s(\eta) < \rho\eta$ for sufficiently large $\eta$; the same holds for $g_{s,i}(\eta)$.*
5. *$g_{s,i}$ and $g_s(\eta)$ have unique positive fixed points $\eta_i^* = g_{s,i}(\eta_i^*)$ and $\eta^* = g_s(\eta^*)$, respectively, and $\eta^* = \max_i \eta_i^*$.*
6. *For $\eta \in (0, \eta^*)$, we have $g_s(\eta) > \eta$, and for $\eta \in (\eta^*, \infty)$, we have $g_s(\eta) < \eta$; the same statement holds also for $g_{s,i}(\eta)$.*
7. *For any $\epsilon > 0$, there exists $\rho_1(\epsilon) > 1$ such that $g_s(\eta) > \rho_1(\epsilon)\eta$ as long as $0 < \eta \leq (1 - \epsilon)\eta^*$, and there exists $\rho_2(\epsilon) < 1$ such that $g_s(\eta) < \rho_2(\epsilon)\eta$ as long as $\eta > (1 + \epsilon)\eta^*$.*

The same properties in Proposition 5 hold for the functions defined below:

$$h_{s,i}(\eta) = \min_{P_i} s\eta \left( \max_j \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 \right) + \sum_{l=1}^{p} \|P_i^l\|_2, \text{ and} \tag{50}$$
$$h_s(\eta) = \max_i h_{s,i}(\eta). \tag{51}$$

**Proposition 6** *The functions $h_{s,i}(\eta)$ and $h_s(\eta)$ have the following properties:*

1. *$h_{s,i}(\eta)$ and $h_s(\eta)$ are continuous in $\eta$.*
2. *$h_{s,i}(\eta)$ and $h_s(\eta)$ are strictly increasing in $\eta$.*
3. *$h_{s,i}(\eta)$ is concave for every $i$.*
4. *$h_s(0) = 0$, $h_s(\eta) \geq s\eta > \eta$ for sufficiently small $\eta > 0$, and there exists $\rho < 1$ such that $h_s(\eta) < \rho\eta$ for sufficiently large $\eta$; the same holds for $h_{s,i}(\eta)$.*

5. $h_{s,i}$ and $h_s(\eta)$ have unique positive fixed points $\eta_i^* = h_{s,i}(\eta_i^*)$ and $\eta^* = h_s(\eta^*)$, respectively, and $\eta^* = \max_i \eta_i^*$.
6. For $\eta \in (0, \eta^*)$, we have $h_s(\eta) > \eta$, and for $\eta \in (\eta^*, \infty)$, we have $h_s(\eta) < \eta$; the same statement holds also for $h_{s,i}(\eta)$.
7. For any $\epsilon > 0$, there exists $\rho_1(\epsilon) > 1$ such that $h_s(\eta) > \rho_1(\epsilon)\eta$ as long as $0 < \eta \leq (1 - \epsilon)\eta^*$, and there exists $\rho_2(\epsilon) < 1$ such that $h_s(\eta) < \rho_2(\epsilon)\eta$ as long as $\eta > (1 + \epsilon)\eta^*$.

An immediate consequence of Propositions 5 and 6 is the following:

**Theorem 4** *Suppose $\eta^*$ is the unique fixed point of $g_s(\eta)$ ($h_s(\eta)$, resp.), then we have*

$$\eta^* \geq \frac{1}{\omega_2(A, s)} \left( \frac{1}{\omega_{b\infty}(A^T A, s)}, \ resp. \right). \tag{52}$$

Proposition 5 implies three ways to compute the fixed point $\eta^*$ for $g_s(\eta)$. The same discussion is also valid for $h_s(\eta)$.

1. **Naive Fixed Point Iteration:** Property 7) of Proposition 5 suggests that the fixed point iteration

$$\eta_{t+1} = g_s(\eta_t), t = 0, 1, \ldots \tag{53}$$

   starting from any initial point $\eta_0$ converges to $\eta^*$, no matter whether $\eta_0 < \eta^*$ or $\eta_0 > \eta^*$. The algorithm can be made more efficient in the case $\eta_0 < \eta^*$. More specifically, since $g_s(\eta) = \max_i g_{s,i}(\eta)$, at each fixed point iteration, we set $\eta_{t+1}$ to be the first $g_{s,i}(\eta_t)$ that is greater than $\eta_t + \epsilon$, with $\epsilon$ being some tolerance parameter. If for all $i$, $g_{s,i}(\eta_t) < \eta_t + \epsilon$, then $g_s(\eta_t) = \max_i g_{s,i}(\eta_t) < \eta_t + \epsilon$, which indicates that the optimal function value can not be improved greatly and the algorithm should terminate. In most cases, to get $\eta_{t+1}$, we need to solve only one optimization problem, $\min_{P_i} s\eta \left( \max_j \|\delta_{ij} I_n - P_i^T Q_j\|_2 \right) + \|P_i\|_2$, instead of solving for $p$. This is in contrast to the case where $\eta_0 > \eta^*$, because in the latter case we must compute all $g_{s,i}(\eta_t)$ to update $\eta_{t+1} = \max_i g_{s,i}(\eta_t)$. An update based on a single $g_{s,i}(\eta_t)$ might generate a value smaller than $\eta^*$.
   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 $\eta_t$ and $\eta^*$. This disadvantage can be remedied by starting from both below and above $\eta^*$. The distance between corresponding terms in the two generated sequences is an indication of the gap to the fixed point $\eta^*$. However, the resulting algorithm is very slow, especially when updating $\eta_{t+1}$ from above $\eta^*$. Second, the iteration process is slow, especially when close to the fixed point $\eta^*$, because $\rho_1(\epsilon)$ and $\rho_2(\epsilon)$ in 7) of Proposition 5 are close to 1.
2. **Bisection:** The bisection approach is motivated by property 6) of Proposition 5. Starting from an initial interval $(\eta_L, \eta_U)$ that contains $\eta^*$, we compute $g_s(\eta_M)$ with $\eta_M = (\eta_L + \eta_U)/2$. As a consequence of property 6), $g_s(\eta_M) > \eta_M$ implies $g_s(\eta_M) < \eta^*$, and we set $\eta_L = g_s(\eta_M)$; $g_s(\eta_M) < \eta_M$

implies $g_s(\eta_{\mathrm{M}}) > \eta^*$, and we set $\eta_{\mathrm{U}} = g_s(\eta_{\mathrm{M}})$. The bisection process can also be accelerated by setting $\eta_{\mathrm{L}} = g_{s,i}(\eta_{\mathrm{M}})$ for the first $g_{s,i}(\eta_{\mathrm{M}})$ greater than $\eta_{\mathrm{M}}$. 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 $\eta_{\mathrm{M}}$ and $\eta^*$, resulting an accurate stopping criterion. However, if the initial $\eta_{\mathrm{U}}$ is too much larger than $\eta^*$, the majority of $g_s(\eta_{\mathrm{M}})$ would turn out to be less than $\eta^*$. The verification of $g_s(\eta_{\mathrm{M}}) < \eta_{\mathrm{M}}$ requires solving $p$ semidefinite programs, greatly degrading the algorithm's performance.

3. **Fixed Point Iteration + Bisection:** The third approach combines the advantages of the bisection method and the fixed point iteration method, at the level of $g_{s,i}(\eta)$. This method relies on the representations $g_s(\eta) = \max_i g_{s,i}(\eta)$ and $\eta^* = \max_i \eta_i^*$.
Starting from an initial interval $(\eta_{\mathrm{L}0}, \eta_{\mathrm{U}})$ and the index set $\mathcal{I}_0 = \{1, \ldots, p\}$, we pick any $i_0 \in \mathcal{I}_0$ and use the (accelerated) bisection method with starting interval $(\eta_{\mathrm{L}0}, \eta_{\mathrm{U}})$ to find the positive fixed point $\eta_{i_0}^*$ of $g_{s,i_0}(\eta)$. For any $i \in \mathcal{I}_0/i_0$, $g_{s,i}(\eta_{i_0}^*) \leq \eta_{i_0}^*$ implies that the fixed point $\eta_i^*$ of $g_{s,i}(\eta)$ is less than or equal to $\eta_{i_0}^*$ according to the continuity of $g_{s,i}(\eta)$ and the uniqueness of its positive fixed point. As a consequence, we remove this $i$ from the index set $\mathcal{I}_0$. We denote $\mathcal{I}_1$ as the index set after all such $i$s are removed, i.e., $\mathcal{I}_1 = \mathcal{I}_0/\{i : g_{s,i}(\eta_{i_0}^*) \leq \eta_{i_0}^*\}$. We also set $\eta_{L1} = \eta_{i_0}^*$ as $\eta^* \geq \eta_{i_0}^*$. Next we test the $i_1 \in \mathcal{I}_1$ with the *largest* $g_{s,i}(\eta_{i_0}^*)$ and construct $\mathcal{I}_2$ and $\eta_{\mathrm{L}2}$ in a similar manner. We repeat the process until the index set $\mathcal{I}_t$ is empty. The $\eta_i^*$ found at the last step is the maximal $\eta_i^*$, which is equal to $\eta^*$.

## 6 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 error bounds based on $\omega_2(A, s)$ with the bounds based on the block RIP [7]. 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 length.

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 in the latter case are computed using the algorithms developed in [20, 21]. We note in Table 1 that for the same sensing matrix $A$, both $s_*$ and $k_*$ are smaller when the block-sparse structure is taken into account than when it is not taken into account. 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.

**Table 1** 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 | |
|---|---|---|---|---|---|
| | $s_*$ | $k_*$ | $nk_*$ | $s_*$ | $k_*$ |
| 72 | 3.96 | 1 | 4 | 6.12 | 3 |
| 96 | 4.87 | 2 | 8 | 7.55 | 3 |
| 120 | 5.94 | 2 | 8 | 9.54 | 4 |
| 144 | 7.14 | 3 | 12 | 11.96 | 5 |
| 168 | 8.60 | 4 | 16 | 14.66 | 7 |
| 192 | 11.02 | 5 | 20 | 18.41 | 9 |

In the next set of experiments, we compare the computation times for the three implementation methods discussed at the end of Section 5.4. 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 $\eta_{\mathrm{L}}$ and upper bound $\eta_{\mathrm{U}}$ 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 RIP. Recall that from Corollary 2, we have for the BS-BP

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{2\sqrt{2k}}{\omega_2(A, 2k)}\varepsilon. \tag{54}$$

For comparison, the block RIP bounds is

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{1 + \delta_{2k}(A)}}{1 - (1 + \sqrt{2})\delta_{2k}(A)}\varepsilon, \tag{55}$$

assuming the block RIP $\delta_{2k}(A) < \sqrt{2} - 1$ [7]. 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 2, 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 (55) does not apply at all. In contrast, the $\omega_2$ based bound (54) is valid as long as $k \leq k_*$. In Table (3) we show the $\omega_2$ based bound (54).

**Table 2** $\omega_2(A, 2k)$ and $\delta_{2k}(A)$ computed for a Gaussian matrix $A \in \mathbb{R}^{m \times np}$ with $n = 4$ and $p = 60$.

| | $m$ | 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 |
| 1 | $\omega_2(A, 2k)$ | 0.45 | 0.53 | 0.57 | 0.62 | 0.65 | 0.67 |
| | $\delta_{2k}(A)$ | 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 |
| | $\delta_{2k}(A)$ | | 1.08 | 0.98 | 0.96 | 0.84 | 0.75 |
| 3 | $\omega_2(A, 2k)$ | —— | —— | —— | 0.11 | 0.18 | 0.25 |
| | $\delta_{2k}(A)$ | | | | 1.12 | 1.01 | 0.93 |
| 4 | $\omega_2(A, 2k)$ | —— | —— | —— | —— | 0.02 | 0.12 |
| | $\delta_{2k}(A)$ | | | | | 1.26 | 1.07 |
| 5 | $\omega_2(A, 2k)$ | —— | —— | —— | —— | —— | 0.03 |
| | $\delta_{2k}(A)$ | | | | | | 1.28 |

**Table 3** 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 2.

| | $m$ | 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 |
| 1 | $\omega_2$ 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 |

## 7 Conclusions

In this paper, we analyzed the performance of convex block-sparse signal recovery algorithms using the block-$\ell_{\mathrm{b}\infty}$ norm of the errors as a performance

criterion. We expressed other popular performance criteria in terms of the block-$\ell_{\mathrm{b}\infty}$ norm. A family of goodness measures of the sensing matrices was defined using optimization procedures. We used these goodness measures to derive upper bounds on the block-$\ell_{\mathrm{b}\infty}$ norms of the reconstruction 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 goodness measures. We expect that these goodness 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.

## 8 Appendix: Proofs

### 8.1 Proof of Proposition 1

*Proof* Suppose $S = \mathrm{bsupp}(\boldsymbol{x})$ and $|S| = \|\boldsymbol{x}\|_{\mathrm{b}0} = k$. Define the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$.

We first prove the proposition for the BS-BP and the BS-DS. The fact that $\|\hat{\boldsymbol{x}}\|_{\mathrm{b}1} = \|\boldsymbol{x} + \boldsymbol{h}\|_{\mathrm{b}1}$ is the minimum among all $\boldsymbol{z}$s satisfying the constraints in (5) and (6), 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, implies that $\|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1}$ cannot be very large. To see this, note that

$$
\begin{aligned}
\|\boldsymbol{x}\|_{\mathrm{b}1} &\geq \|\boldsymbol{x} + \boldsymbol{h}\|_{\mathrm{b}1} \\
&= \sum_{i \in S} \|\boldsymbol{x}_i + \boldsymbol{h}_i\|_2 + \sum_{i \in S^c} \|\boldsymbol{x}_i + \boldsymbol{h}_i\|_2 \\
&\geq \sum_{i \in S} \|\boldsymbol{x}_i\|_2 - \sum_{i \in S} \|\boldsymbol{h}_i\|_2 + \sum_{i \in S^c} \|\boldsymbol{h}_i\|_2 \\
&= \|\boldsymbol{x}_S\|_{\mathrm{b}1} - \|\boldsymbol{h}_S\|_{\mathrm{b}1} + \|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1} \\
&= \|\boldsymbol{x}\|_{\mathrm{b}1} - \|\boldsymbol{h}_S\|_{\mathrm{b}1} + \|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1}. \quad (56)
\end{aligned}
$$

Therefore, we obtain $\|\boldsymbol{h}_S\|_{\mathrm{b}1} \geq \|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1}$, which leads to

$$
2\|\boldsymbol{h}_S\|_{\mathrm{b}1} \geq \|\boldsymbol{h}_S\|_{\mathrm{b}1} + \|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1} = \|\boldsymbol{h}\|_{\mathrm{b}1}. \quad (57)
$$

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

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

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

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

Using the Cauchy-Swcharz type inequality, we get

$$
\begin{aligned}
\mu\|\hat{\boldsymbol{x}}\|_{\mathrm{b}1} &\leq \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_{\mathrm{b}1}\|A^T \boldsymbol{w}\|_{\mathrm{b}\infty} + \mu\|\boldsymbol{x}\|_{\mathrm{b}1} \\
&= \kappa\mu\|\boldsymbol{h}\|_{\mathrm{b}1} + \mu\|\boldsymbol{x}\|_{\mathrm{b}1},
\end{aligned}
$$

which leads to

$$
\|\hat{\boldsymbol{x}}\|_{\mathrm{b}1} \leq \kappa\|\boldsymbol{h}\|_{\mathrm{b}1} + \|\boldsymbol{x}\|_{\mathrm{b}1}.
$$

Therefore, similar to the argument in (56), we have

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

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

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

Therefore, similar to (57), we obtain

$$
\begin{aligned}
\frac{2}{1 - \kappa}\|\boldsymbol{h}_S\|_{\mathrm{b}1} &= \frac{1 + \kappa}{1 - \kappa}\|\boldsymbol{h}_S\|_{\mathrm{b}1} + \frac{1 - \kappa}{1 - \kappa}\|\boldsymbol{h}_S\|_{\mathrm{b}1} \\
&\geq \frac{1 + \kappa}{1 - \kappa}\frac{1 - \kappa}{1 + \kappa}\|\boldsymbol{h}_{S^c}\|_{\mathrm{b}1} + \frac{1 - \kappa}{1 - \kappa}\|\boldsymbol{h}_S\|_{\mathrm{b}1} \\
&= \|\boldsymbol{h}\|_{\mathrm{b}1}. \tag{58}
\end{aligned}
$$

$\square$

8.2 Proof of Corollary 1

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

$$\|\boldsymbol{h}\|_{\mathrm{b}1} \leq c\|\boldsymbol{h}_S\|_{\mathrm{b}1} \leq ck\|\boldsymbol{h}\|_{\mathrm{b}\infty}. \tag{59}$$

To prove (13), we note

$$\begin{aligned}
\frac{\|\boldsymbol{h}\|_2^2}{\|\boldsymbol{h}\|_{\mathrm{b}\infty}^2} &= \sum_{i=1}^p \left( \frac{\|\boldsymbol{h}_i\|_2}{\|\boldsymbol{h}\|_{\mathrm{b}\infty}} \right)^2 \\
&\leq \sum_{i=1}^p \left( \frac{\|\boldsymbol{h}_i\|_2}{\|\boldsymbol{h}\|_{\mathrm{b}\infty}} \right) \\
&= \frac{\|\boldsymbol{h}\|_{\mathrm{b}1}}{\|\boldsymbol{h}\|_{\mathrm{b}\infty}} \\
&\leq ck.
\end{aligned} \tag{60}$$

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

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

$$\|\hat{\boldsymbol{x}}_i\|_2 = \|\boldsymbol{x}_i + \boldsymbol{h}_i\|_2 \geq \|\boldsymbol{x}_i\|_2 - \|\boldsymbol{h}_i\|_2 > \beta - \beta/2 = \beta/2; \tag{61}$$

and for $i \notin S$,

$$\|\hat{\boldsymbol{x}}_i\|_2 = \|\boldsymbol{x}_i + \boldsymbol{h}_i\|_2 = \|\boldsymbol{h}_i\|_2 < \beta/2. \tag{62}$$

$\square$

8.3 Proof of Theorem 1

*Proof* Observe that for the BS-BP

$$\begin{aligned}
\|A(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_2 &\leq \|\boldsymbol{y} - A\hat{\boldsymbol{x}}\|_2 + \|\boldsymbol{y} - A\boldsymbol{x}\|_2 \\
&\leq \varepsilon + \|A\boldsymbol{w}\|_2 \\
&\leq 2\varepsilon,
\end{aligned} \tag{63}$$

and similarly,

$$\|A^T A(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_{\mathrm{b}\infty} \leq 2\mu \tag{64}$$

for the BS-DS, and

$$\|A^T A(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_{\mathrm{b}\infty} \leq (1 + \kappa)\mu \tag{65}$$

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

$$\|A^T(A\hat{\boldsymbol{x}} - \boldsymbol{y})\|_{\mathrm{b}\infty} \leq \mu, \tag{66}$$

a consequence of the optimality condition

$$A^T(A\hat{\boldsymbol{x}} - \boldsymbol{y}) \in \mu\partial\|\hat{\boldsymbol{x}}\|_{\mathrm{b1}} \tag{67}$$

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

The conclusions of Theorem 1 follow from equation (12) and Definition 1. $\qquad\square$

8.4 Proof of Lemma 1

*Proof* For any $\boldsymbol{z}$ such that $\|\boldsymbol{z}\|_{\mathrm{b\infty}} = 1$ and $\|\boldsymbol{z}\|_{\mathrm{b1}} \leq s$, we have

$$\begin{aligned}
\boldsymbol{z}A^TA\boldsymbol{z} &= \langle \boldsymbol{z}, A^TA\boldsymbol{z} \rangle \\
&\leq \|\boldsymbol{z}\|_{\mathrm{b1}}\|A^TA\boldsymbol{z}\|_{\mathrm{b\infty}} \\
&\leq s\|A^TA\boldsymbol{z}\|_{\mathrm{b\infty}}.
\end{aligned} \tag{68}$$

Taking the minima of both sides of (68) over $\{\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b\infty}} = 1, \|\boldsymbol{z}\|_1 \leq s\}$ yields

$$\omega_2^2(A, s) \leq s\omega_{\mathrm{b\infty}}(A^TA, s). \tag{69}$$

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

$$\{\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b1}}/\|\boldsymbol{z}\|_{\mathrm{b\infty}} \leq s\} \subseteqq \{\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b1}}/\|\boldsymbol{z}\|_2 \leq s\}. \tag{70}$$

As a consequence, we have

$$\begin{aligned}
\omega_2(A, s) &= \min_{\|\boldsymbol{z}\|_{\mathrm{b1}}/\|\boldsymbol{z}\|_{\mathrm{b\infty}} \leq s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}\frac{\|\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_{\mathrm{b\infty}}} \\
&\geq \min_{\|\boldsymbol{z}\|_{\mathrm{b1}}/\|\boldsymbol{z}\|_{\mathrm{b\infty}} \leq s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \\
&\geq \min_{\|\boldsymbol{z}\|_{\mathrm{b1}}/\|\boldsymbol{z}\|_2 \leq s} \frac{\|A\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \\
&= \rho_{s^2}(A),
\end{aligned} \tag{71}$$

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

8.5 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

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

The $\psi_2$ norm is closely related to the subgaussian constant $L$. We actually can equivalently define that a random vector $\boldsymbol{X} \in \mathbb{R}^{np}$ is isotropic and subgaussian with constant $L$ if $\mathbb{E}|\langle \boldsymbol{X}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\|\langle \boldsymbol{X}, \boldsymbol{u}\rangle\|_{\psi_2} \leq L\|\boldsymbol{u}\|_2$ for any $\boldsymbol{u} \in \mathbb{R}^{np}$ [20].

We use the notation $\ell_*(\mathcal{H}) = \mathbb{E}\sup_{\boldsymbol{u}\in\mathcal{H}}\langle \boldsymbol{g}, \boldsymbol{u}\rangle$ with $\boldsymbol{g} \sim \mathcal{N}(0, I_{np})$, *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 now cite a result on the behavior of empirical processes [13, 20]:

**Theorem 5** [13, 20] *Let $\{\boldsymbol{a}, \boldsymbol{a}_i, i = 1, \ldots, m\} \subset \mathbb{R}^{np}$ be i.i.d. isotropic and subgaussian random vectors. $\mathcal{H}$ is a subset of the unit sphere of $\mathbb{R}^{np}$, and $\mathcal{F} = \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot\rangle : \boldsymbol{u} \in \mathcal{H}\}$. Suppose $\mathrm{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) \stackrel{\mathrm{def}}{=} \max_{f,g\in\mathcal{F}}\|f - g\|_{\psi_2} = \alpha$. Then there exist absolute constants $c_1, c_2, c_3$ such that for any $\epsilon > 0$ and $m \geq 1$ satisfying*

$$m \geq c_1\frac{\alpha^2\ell_*^2(\mathcal{H})}{\epsilon^2}, \tag{73}$$

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

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^{m}f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right| \leq \epsilon. \tag{74}$$

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

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^{m}f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right|$$
$$\leq c_3\max\left\{\alpha\frac{\ell_*(\mathcal{H})}{\sqrt{m}}, \frac{\ell_*^2(\mathcal{H})}{m}\right\}. \tag{75}$$

With these preparations, we proceed to the proof of Theorem 2:

*Proof (Proof of Theorem 2)* We apply Theorem 5 to estimate the block $\ell_1$-CMSV. According to the assumptions of Theorem 2, the rows $\{\boldsymbol{a}_i\}_{i=1}^m$ of $\sqrt{m}A$ are *i.i.d.* isotropic and subguassian random vectors with constant $L$. Consider the set $\mathcal{H} = \{\boldsymbol{u} \in \mathbb{R}^{np} : \|\boldsymbol{u}\|_2^2 = 1, \|\boldsymbol{u}\|_{\mathrm{b1}}^2 \leq s\}$ and the function set $\mathcal{F} = \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot\rangle : \boldsymbol{u} \in \mathcal{H}\}$. Clearly both $\mathcal{H}$ and $\mathcal{F}$ are symmetric. The diameter of $\mathcal{F}$ satisfies

$$\alpha = \mathrm{diam}(\mathcal{F}, \|\cdot\|_{\psi_2})$$
$$\leq 2\sup_{\boldsymbol{u}\in\mathcal{H}}\|\langle \boldsymbol{u}, \boldsymbol{a}\rangle\|_{\psi_2} = 2L. \tag{76}$$

Since $\mathbb{E} f^2(\boldsymbol{a}) = \mathbb{E} \langle \boldsymbol{u}, \boldsymbol{a} \rangle^2 = \|\boldsymbol{u}\|_2^2 = 1$ when $\boldsymbol{a}$ follows the same distribution as $\{\boldsymbol{a}_i\}_{i=1}^m$, we observe that for any $\epsilon \in (0,1)$

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

is a consequence of

$$\sup_{\boldsymbol{u}\in\mathcal{H}} \left| \frac{1}{m} \boldsymbol{u}^T (\sqrt{m}A)^T (\sqrt{m}A) \boldsymbol{u} - 1 \right|$$

$$= \sup_{\boldsymbol{u}\in\mathcal{H}} \left| \frac{1}{m} \sum_{i=1}^m f_{\boldsymbol{u}}^2(\boldsymbol{a}_i) - \mathbb{E} f_{\boldsymbol{u}}^2(\boldsymbol{a}) \right| \le \epsilon. \tag{78}$$

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

$$\begin{aligned}
\ell_*(\mathcal{H}) &= \mathbb{E} \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|_2=1,\|\boldsymbol{u}\|_{\mathrm{b1}}^2\le s} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \\
&\le \mathbb{E}\|\boldsymbol{u}\|_{\mathrm{b1}}\|\boldsymbol{g}\|_{\mathrm{b}\infty} \\
&\le \sqrt{s}\, \mathbb{E}\|\boldsymbol{g}\|_{\mathrm{b}\infty}.
\end{aligned} \tag{79}$$

The quantity $\mathbb{E}\|\boldsymbol{g}\|_{\mathrm{b}\infty}$ can be bounded using Slepian's inequality [10, Section 3.3]. We rearrange the vector $\boldsymbol{g} \in \mathbb{R}^{np}$ into a matrix $G \in \mathbb{R}^{n\times p}$ such that the vectorization $\mathrm{vec}(G) = \boldsymbol{g}$. Clearly, we have $\|\boldsymbol{g}\|_{\mathrm{b}\infty} = \|G\|_{1,2}$, 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 $\|G\|_{1,2} = \max_{\boldsymbol{v}\in S^{n-1}, \boldsymbol{w}\in T^{p-1}} \langle G\boldsymbol{v}, \boldsymbol{w} \rangle$, we define the Gaussian process $X_{\boldsymbol{v},\boldsymbol{w}} = \langle G\boldsymbol{v}, \boldsymbol{w} \rangle$ indexed by $(\boldsymbol{v}, \boldsymbol{w}) \in S^{n-1} \times T^{p-1}$. Here $S^{n-1} = \{\boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{v}\|_2 = 1\}$ and $T^{p-1} = \{\boldsymbol{w} \in \mathbb{R}^p : \|\boldsymbol{w}\|_1 = 1\}$. We compare $X_{\boldsymbol{v},\boldsymbol{w}}$ with another Gaussian process $Y_{\boldsymbol{v},\boldsymbol{w}} = \langle \boldsymbol{\xi}, \boldsymbol{v} \rangle + \langle \boldsymbol{\zeta}, \boldsymbol{w} \rangle, (\boldsymbol{v}, \boldsymbol{w}) \in S^{n-1} \times T^{p-1}$, where $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{I}_n)$ and $\boldsymbol{\zeta} \sim \mathcal{N}(0, \mathrm{I}_p)$. The Gaussian processes $X_{\boldsymbol{v},\boldsymbol{w}}$ and $Y_{\boldsymbol{v},\boldsymbol{w}}$ satisfy the conditions for Slepian's inequality (See the proof of [23, Theorem 32, page 23]). Therefore, we have

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{g}\|_{\mathrm{b}\infty} &= \mathbb{E} \max_{(\boldsymbol{v},\boldsymbol{w})\in S^{n-1}\times T^{p-1}} X_{\boldsymbol{v},\boldsymbol{w}} \le \mathbb{E} \max_{(\boldsymbol{v},\boldsymbol{w})\in S^{n-1}\times T^{p-1}} Y_{\boldsymbol{v},\boldsymbol{w}} \\
&= \mathbb{E} \max_{\boldsymbol{v}\in S^{n-1}} \langle \boldsymbol{\xi}, \boldsymbol{v} \rangle + \mathbb{E} \max_{\boldsymbol{w}\in T^{p-1}} \langle \boldsymbol{\zeta}, \boldsymbol{w} \rangle \\
&= \mathbb{E}\|\boldsymbol{\xi}\|_2 + \mathbb{E}\|\boldsymbol{\zeta}\|_\infty \\
&\le \sqrt{n} + \sqrt{\log p}.
\end{aligned} \tag{80}$$

Here we have used

$$\mathbb{E}\|\boldsymbol{\xi}\|_2 \le \sqrt{\mathbb{E}\|\boldsymbol{\xi}\|_2^2} = \sqrt{n} \tag{81}$$

due to Jensen's inequality, and

$$\mathbb{E}\|\boldsymbol{\zeta}\|_\infty = \mathbb{E} \max_i \boldsymbol{\zeta}_i \le \sqrt{\log p}, \tag{82}$$

a fact given by [10, Equation 3.13, page 79].

The conclusion of Theorem 2 then follows from (79), (80), and suitable choice of $c_1$. $\qquad\square$

8.6 Proof of Proposition 2

*Proof* We rewrite the optimization (33) as

$$\frac{1}{s^*} = \max_{\boldsymbol{z}} \|\boldsymbol{z}\|_{\mathrm{b}\infty} \text{ s.t. } Q\boldsymbol{z} = 0, \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1. \tag{83}$$

Note that in (83), 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 (83). Define a matrix variable $P$ of the same size as $Q$. Since the dual norm of $\|\cdot\|_{\mathrm{b}\infty}$ is $\|\cdot\|_{\mathrm{b}1}$, we have

$$
\begin{aligned}
&\max_{\boldsymbol{z}} \left\{ \|\boldsymbol{z}\|_{\mathrm{b}\infty} : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, Q\boldsymbol{z} = 0 \right\} \\
&= \max_{\boldsymbol{u},\boldsymbol{z}} \left\{ \boldsymbol{u}^T \boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1, Q\boldsymbol{z} = 0 \right\} \\
&= \max_{\boldsymbol{u},\boldsymbol{z}} \left\{ \boldsymbol{u}^T (\boldsymbol{z} - P^T Q \boldsymbol{z}) : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1, Q\boldsymbol{z} = 0 \right\} \\
&\le \max_{\boldsymbol{u},\boldsymbol{z}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1 \right\}. \tag{84}
\end{aligned}
$$

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

$$
\begin{aligned}
&\max_{\boldsymbol{u},\boldsymbol{z}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1 \right\} \\
&= \max_{j,\boldsymbol{v},\boldsymbol{u}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)(\mathbf{e}_i^p \otimes \boldsymbol{v}) : \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{v}\|_2 \le 1 \right\} \\
&= \max_{j} \max_{\boldsymbol{v},\boldsymbol{u}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)_j \boldsymbol{v} : \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{v}\|_2 \le 1 \right\} \\
&= \max_{j} \max_{\boldsymbol{u}} \left\{ \|(\mathrm{I}_{np} - P^T Q)_j^T \boldsymbol{u}\|_2 : \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1 \right\}, \tag{85}
\end{aligned}
$$

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

Applying the same argument to the unit ball $\{\boldsymbol{u} : \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1\}$ and the convex function $\|(\mathrm{I}_{np} - P^T Q)_j^T \boldsymbol{u}\|_2$, we obtain

$$
\begin{aligned}
&\max_{\boldsymbol{u},\boldsymbol{z}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \le 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \le 1 \right\} \\
&= \max_{i,j} \|(\mathrm{I}_{np} - P^T Q)_{i,j}\|_2. \tag{86}
\end{aligned}
$$

Here $(\mathrm{I}_{np} - P^T Q)_{i,j}$ is the submatrix of $\mathrm{I}_{np} - P^T Q$ formed by the $((i-1)n+1)$th to $in$th rows and the $((j-1)n+1)$th to $jn$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 $\max_{i,j} \|(\mathrm{I}_{np} - P^T Q)_{i,j}\|_2$ with respect to $P$:

$$
\begin{aligned}
1/s^* &= \max_{\boldsymbol{z}} \{\|\boldsymbol{z}\|_{\mathrm{b}\infty} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq 1, Q\boldsymbol{z} = 0\} \\
&\leq \min_{P} \max_{i,j} \|(\mathrm{I}_{np} - P^T Q)_{i,j}\|_2 \\
&\overset{\mathrm{def}}{=} 1/s_*.
\end{aligned} \tag{87}
$$

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

$$
(\mathrm{I}_{np} - P^T Q)_{i,j} = \delta_{ij} \mathrm{I}_n - P_i^T Q_j, \tag{88}
$$

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

$$
\begin{aligned}
&\min_{P} \max_{i,j} \|(\mathrm{I}_{np} - P^T Q)_{i,j}\|_2 \\
&= \min_{P_1,\ldots,P_p} \max_{i} \max_{j} \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2 \\
&= \max_{i} \min_{P_i} \max_{j} \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2.
\end{aligned} \tag{89}
$$

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

8.7 Proof of Proposition 3

*Proof* 1. Since in the optimization problem defining $f_s(\eta)$, the objective function $\|\boldsymbol{z}\|_{\mathrm{b}\infty}$ is continuous, and the constraint correspondence

$$
\begin{aligned}
C(\eta) &: [0, \infty) \twoheadrightarrow \mathbb{R}^{np} \\
&\eta \mapsto \{\boldsymbol{z} : \|Q\boldsymbol{z}\|_\diamond \leq 1, \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta\}
\end{aligned} \tag{90}
$$

is compact-valued and continuous (both upper and lower hemicontinuous), according to Berge's Maximum Theorem [1] the optimal value function $f_s(\eta)$ is continuous.
2. The monotone (non-strict) increasing property is obvious as increasing $\eta$ enlarges the region over which the maximization is taken. We now show the strict increasing property. Suppose $0 < \eta_1 < \eta_2$, and $f_s(\eta_1)$ is achieved by $\boldsymbol{z}_1^* \neq 0$, namely, $f_s(\eta_1) = \|\boldsymbol{z}_1^*\|_{\mathrm{b}\infty}$, $\|Q\boldsymbol{z}_1^*\|_\diamond \leq 1$, and $\|\boldsymbol{z}_1^*\|_{\mathrm{b}1} \leq s\eta_1$. If $\|Q\boldsymbol{z}_1^*\|_\diamond < 1$ (this implies $\|\boldsymbol{z}_1^*\|_{\mathrm{b}1} = s\eta_1$), we define $\boldsymbol{z}_2 = c\boldsymbol{z}_1^*$ with $c = \min(1/\|Q\boldsymbol{z}_1^*\|_\diamond, s\eta_2/\|\boldsymbol{z}_1^*\|_{\mathrm{b}1}) > 1$. We then have $\|Q\boldsymbol{z}_2\|_\diamond \leq 1$, $\|\boldsymbol{z}_2\|_{\mathrm{b}1} \leq s\eta_2$, and $f_s(\eta_2) \geq \|\boldsymbol{z}_2\|_{\mathrm{b}\infty} = c\|\boldsymbol{z}_1^*\|_{\mathrm{b}\infty} > f_s(\eta_1)$.
Consider the remaining case that $\|Q\boldsymbol{z}_1^*\|_\diamond = 1$. Without loss of generality, suppose $\|\boldsymbol{z}_1^*\|_{\mathrm{b}\infty} = \max_{1 \leq j \leq p} \|\boldsymbol{z}_{1j}^*\|_2$ is achieved by the block $\boldsymbol{z}_{11}^* \neq 0$. Since $Q_1 \boldsymbol{z}_{11}^*$ is linearly dependent with the columns of $\{Q_j, j = 2, \ldots, p\}$ ($m$ is much less than $(p-1)n$), there exist $\{\boldsymbol{\alpha}_j \in \mathbb{R}^n\}_{j=2}^p$ such that

$Q_1 z_{11}^* + \sum_{j=2}^p Q_j \alpha_j = 0$. Define $\boldsymbol{\alpha} = \left[ z_{11}^{*T}, \boldsymbol{\alpha}_2^T, \cdots, \boldsymbol{\alpha}_p^T \right] \in \mathbb{R}^{np}$ satisfying $Q\boldsymbol{\alpha} = 0$ and $z_2 = z_1^* + c\boldsymbol{\alpha}$ for $c > 0$ sufficiently small such that $\|z_2\|_{\mathrm{b}1} \le \|z_1^*\|_{\mathrm{b}1} + c\|\boldsymbol{\alpha}\|_{\mathrm{b}1} \le s\eta_1 + c\|\boldsymbol{\alpha}\|_{\mathrm{b}1} \le s\eta_2$. Clearly, $\|Qz_2\|_\diamond = \|Qz_1^* + cQ\boldsymbol{\alpha}\|_\diamond = \|Qz_1^*\|_\diamond = 1$. As a consequence, we have $f_s(\eta_2) \ge \|z_2\|_{\mathrm{b}\infty} \ge \|z_{21}\|_2 = (1 + c)\|z_{11}^*\|_2 > \|z_1^*\|_{\mathrm{b}\infty} = f_s(\eta_1)$.

The case for $\eta_1 = 0$ is proved by continuity.

3. Next we show $f_s(\eta) > s\eta$ for sufficiently small $\eta > 0$. Take $z$ as the vector whose first element is $s\eta$ and zero otherwise. We have $\|z\|_{\mathrm{b}1} = s\eta$ and $\|z\|_{\mathrm{b}\infty} = s\eta > \eta$ (recall $s \in (1, s^*)$). In addition, when $\eta > 0$ is sufficiently small, we also have $\|Qz\|_\diamond \le 1$. Therefore, for sufficiently small $\eta$, we have $f_s(\eta) \ge s\eta > \eta$.

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

$$f_s(\eta) < \rho_B \eta, \ \forall \ \eta > \eta_B. \tag{91}$$

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 $f_s(\eta) \ge \rho_B \eta$. 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

$$\lim_{k\to\infty} \eta^{(k)} = \infty,$$
$$\lim_{k\to\infty} \rho^{(k)} = 1,$$
$$\rho^{(k)}\eta^{(k)} \le f_s(\eta^{(k)}) = \|z^{(k)}\|_{\mathrm{b}\infty},$$
$$\|Qz^{(k)}\|_\diamond \le 1,$$
$$\|z^{(k)}\|_{\mathrm{b}1} \le s\eta^{(k)}. \tag{92}$$

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 $c\|z_2^{(k)}\| \le \|Qz_2^{(k)}\|_\diamond \le 1$ where $c = \inf_{z:z\perp\mathrm{null}(Q)} \|Qz\|_\diamond / \|z\| > 0$ and $\|\cdot\|$ is any norm. Then $\infty = \lim_{k\to\infty} \|z^{(k)}\|_{\mathrm{b}\infty} \le \lim_{k\to\infty}(\|z_1^{(k)}\|_{\mathrm{b}\infty} + \|z_2^{(k)}\|_{\mathrm{b}\infty})$ implies $\{z_1^{(k)}\}_{k=1}^\infty$ is unbounded. For sufficiently large $k$, we proceed as follows:

$$s^* > \frac{s}{\left(\frac{s}{s^*}\right)^{1/4}} \ge \frac{s\eta^{(k)}}{\rho^{(k)}\eta^{(k)}} \ge \frac{\|z^{(k)}\|_{\mathrm{b}1}}{\|z^{(k)}\|_{\mathrm{b}\infty}}$$
$$\ge \frac{\|z_1^{(k)}\|_{\mathrm{b}1} - \|z_2^{(k)}\|_{\mathrm{b}1}}{\|z_1^{(k)}\|_{\mathrm{b}\infty} + \|z_2^{(k)}\|_{\mathrm{b}\infty}} \ge \left(\frac{s}{s^*}\right)^{1/4} \frac{\|z_1^{(k)}\|_{\mathrm{b}1}}{\|z_1^{(k)}\|_{\mathrm{b}\infty}}, \tag{93}$$

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

$$\frac{\|z_1^{(k)}\|_{\mathrm{b}1}}{\|z_1^{(k)}\|_{\mathrm{b}\infty}} \le \frac{s}{\sqrt{\frac{s}{s^*}}} = \sqrt{ss^*} < s^* \text{ with } Qz_1^{(k)} = 0, \tag{94}$$

which contradicts the definition of $s^*$.

4. Next we show $f_s(\eta)$ has a unique positive fixed point $\eta^*$, which is equal to $\gamma^* \stackrel{\mathrm{def}}{=} 1/\omega_\diamond(Q, s)$. Properties 1) and 3) imply that there must be at least one fixed point.

   To show the uniqueness, we first prove $\gamma^* \geq \eta^*$ for any fixed point $\eta^* = f_s(\eta^*)$. Suppose $\boldsymbol{z}^*$ achieves the optimal value of the optimization defining $f_s(\eta^*)$, *i.e.*,

$$\eta^* = f_s(\eta^*) = \|\boldsymbol{z}^*\|_{\mathrm{b}\infty}, \|Q\boldsymbol{z}^*\|_\diamond \leq 1, \|\boldsymbol{z}^*\|_{\mathrm{b}1} \leq s\eta^*. \tag{95}$$

   Since $\|\boldsymbol{z}^*\|_{\mathrm{b}1}/\|\boldsymbol{z}^*\|_{\mathrm{b}\infty} \leq s\eta^*/\eta^* \leq s$, we have

$$\gamma^* \geq \frac{\|\boldsymbol{z}^*\|_{\mathrm{b}\infty}}{\|Q\boldsymbol{z}^*\|_\diamond} \geq \eta^*. \tag{96}$$

   If $\eta^* < \gamma^*$, we define $\eta_0 = (\eta^* + \gamma^*)/2$ and

$$\boldsymbol{z}^{\mathrm{c}} = \mathrm{argmax}_{\boldsymbol{z}} \frac{s\|\boldsymbol{z}\|_{\mathrm{b}\infty}}{\|\boldsymbol{z}\|_{\mathrm{b}1}} \text{ s.t. } \|Q\boldsymbol{z}\|_\diamond \leq 1, \|\boldsymbol{z}\|_{\mathrm{b}\infty} \geq \eta_0, \text{ and}$$

$$\rho = \frac{s\|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}\infty}}{\|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}1}}. \tag{97}$$

   Suppose $\boldsymbol{z}^{**}$ with $\|Q\boldsymbol{z}^{**}\|_\diamond = 1$ achieves the optimal value of the optimization defining $\gamma^* = 1/\omega_\diamond(Q, s)$. Clearly, $\|\boldsymbol{z}^{**}\|_{\mathrm{b}\infty} = \gamma^* > \eta_0$, which implies $\boldsymbol{z}^{**}$ is a feasible point of the optimization defining $\boldsymbol{z}^{\mathrm{c}}$ and $\rho$. As a consequence, we have

$$\rho \geq \frac{s\|\boldsymbol{z}^{**}\|_{\mathrm{b}\infty}}{\|\boldsymbol{z}^{**}\|_{\mathrm{b}1}} \geq 1. \tag{98}$$

   Actually we will show that $\rho > 1$. If $\|\boldsymbol{z}^{**}\|_{\mathrm{b}1} < s\|\boldsymbol{z}^{**}\|_{\mathrm{b}\infty}$, we are done. If not (*i.e.*, $\|\boldsymbol{z}^{**}\|_{\mathrm{b}1} = s\|\boldsymbol{z}^{**}\|_{\mathrm{b}\infty}$), as illustrated in Figure 1, we consider $\boldsymbol{\xi} = \frac{\eta_0}{\gamma^*}\boldsymbol{z}^{**}$, which satisfies

$$\|Q\boldsymbol{\xi}\|_\diamond \leq \frac{\eta_0}{\gamma^*} < 1, \tag{99}$$

$$\|\boldsymbol{\xi}\|_{\mathrm{b}\infty} = \eta_0, \text{ and} \tag{100}$$

$$\|\boldsymbol{\xi}\|_{\mathrm{b}1} = s\eta_0. \tag{101}$$

   To get $\boldsymbol{\xi}^{\mathrm{n}}$ as shown in Figure 1, pick the block of $\boldsymbol{\xi}$ with smallest non-zero $\ell_2$ norm, and scale that block by a small positive constant less than 1. Because $s > 1$, $\boldsymbol{\xi}$ has more than one non-zero blocks, implying $\|\boldsymbol{\xi}^{\mathrm{n}}\|_{\mathrm{b}\infty}$ will remain the same. If the scaling constant is close enough to 1, $\|Q\boldsymbol{\xi}^{\mathrm{n}}\|_\diamond$ will remain less than 1. But the good news is that $\|\boldsymbol{\xi}^{\mathrm{n}}\|_{\mathrm{b}1}$ decreases, and hence $\rho \geq \frac{s\|\boldsymbol{\xi}^{\mathrm{n}}\|_{\mathrm{b}\infty}}{\|\boldsymbol{\xi}^{\mathrm{n}}\|_{\mathrm{b}1}}$ becomes greater than 1.

   Now we proceed to obtain the contradiction that $f_s(\eta^*) > \eta^*$. If $\|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}1} \leq s \cdot \eta^*$, then it is a feasible point of

$$\max_{\boldsymbol{z}} \|\boldsymbol{z}\|_{\mathrm{b}\infty} \text{ s.t. } \|Q\boldsymbol{z}\|_\diamond \leq 1, \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s \cdot \eta^*. \tag{102}$$
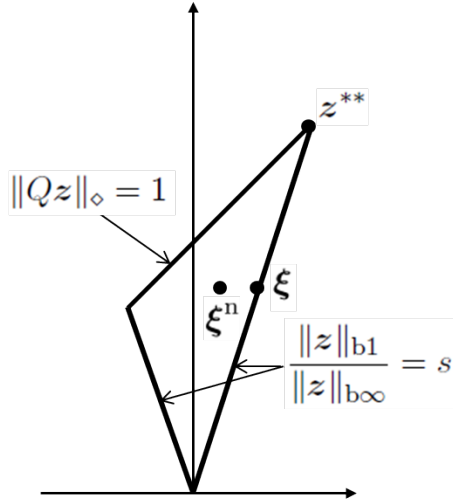
**Fig. 1** Illustration of the proof for $\rho > 1$.

As a consequence, $f_s(\eta^*) \geq \|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}\infty} \geq \eta_0 > \eta^*$, contradicting that $\eta^*$ is a fixed point, and we are done. If this is not the case, *i.e.*, $\|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}1} > s \cdot \eta^*$, we define a new point

$$\boldsymbol{z}^{\mathrm{n}} = \tau \boldsymbol{z}^{\mathrm{c}} \tag{103}$$

with

$$\tau = \frac{s \cdot \eta^*}{\|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}1}} < 1. \tag{104}$$

Note that $\boldsymbol{z}^{\mathrm{n}}$ is a feasible point of the optimization defining $f_s(\eta^*)$ since

$$\|Q\boldsymbol{z}^{\mathrm{n}}\|_{\diamond} = \tau \|Q\boldsymbol{z}^{\mathrm{c}}\|_{\diamond} < 1, \text{ and} \tag{105}$$

$$\|\boldsymbol{z}^{\mathrm{n}}\|_{\mathrm{b}1} = \tau \|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}1} = s \cdot \eta^*. \tag{106}$$

Furthermore, we have

$$\|\boldsymbol{z}^{\mathrm{n}}\|_{\mathrm{b}\infty} = \tau \|\boldsymbol{z}^{\mathrm{c}}\|_{\mathrm{b}\infty} = \rho \eta^*. \tag{107}$$

As a consequence, we obtain

$$f_s(\eta^*) \geq \rho \eta^* > \eta^*. \tag{108}$$

Therefore, for any positive fixed point $\eta^*$, we have $\eta^* = \gamma^*$, *i.e.*, the positive fixed point is unique.

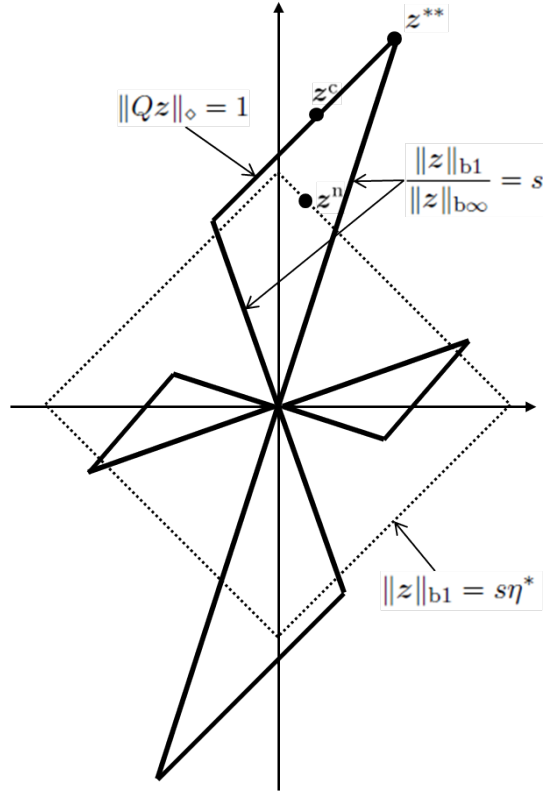5. Property 5) is a consequence of 1), 3), and 4).

**Fig. 2** Illustration of the proof for $f_s(\eta^*) \geq \rho\eta^*$.

6. We demonstrate the existence of $\rho_2(\epsilon)$ only. The existence of $\rho_1(\epsilon)$ can be proved in a similar manner, and hence is omitted.
   We need to show that for fixed $\epsilon > 0$, there exists $\rho(\epsilon) < 1$ such that for any $\eta \geq (1 + \epsilon)\eta^*$ we have

$$f_s(\eta) \leq \rho(\epsilon)\eta. \tag{109}$$

In view of (91), we need to prove the above statement only for $\eta \in [(1 + \epsilon)\eta^*, \eta_B]$. We use contradiction. Suppose for any $\rho \in (0, 1)$ there exists $\eta \in [(1 + \epsilon)\eta^*, \eta_B]$ such that $f_s(\eta) > \rho\eta$. Construct sequences $\{\eta^{(k)}\}_{k=1}^{\infty} \subset [(1 + \epsilon)\eta^*, \eta_B]$ and $\{\rho^{(k)}\}_{k=1}^{\infty} \subset (0, 1)$ with

$$\lim_{k \to \infty} \rho^{(k)} = 1, \text{ and}$$
$$f_s(\eta^{(k)}) > \rho^{(k)}\eta^{(k)}. \tag{110}$$

Due to the compactness of $[(1 + \epsilon)\eta^*, \eta_B]$, there must exist a subsequence $\{\eta^{(k_l)}\}_{l=1}^{\infty}$ of $\{\eta^{(k)}\}_{k=1}^{\infty}$ such that $\lim_{l \to \infty} \eta^{(k_l)} = \eta_{\lim}$ for some $\eta_{\lim} \in [(1 + \epsilon)\eta^*, \eta_B]$. As a consequence of the continuity of $f_s(\eta)$, we have

$$f_s(\eta_{\lim}) = \lim_{l \to \infty} f_s(\eta^{(k_l)}) \geq \lim_{l \to \infty} \rho^{(k_l)} \eta^{(k_l)} = \eta_{\lim}. \qquad (111)$$

Again due to the continuity of $f_s(\eta)$ and the fact that $f_s(\eta) < \eta$ for $\eta > \eta_B$, there exists $\eta_c \in [\eta_{\lim}, \eta_B]$ such that

$$f_s(\eta_c) = \eta_c, \qquad (112)$$

contradicting the uniqueness of the fixed point for $f_s(\eta)$.

The result implies that starting from any initial point below the fixed point, through the iteration $\eta_{t+1} = f_s(\eta_t)$ we can approach an arbitrarily small neighborhood of the fixed point exponentially fast.

□

### 8.8 Proof of Proposition 4

*Proof* Introducing an additional variable $\boldsymbol{v}$ such that $Q\boldsymbol{z} = \boldsymbol{v}$, we have

$$\max_{\boldsymbol{z}} \left\{ \|\boldsymbol{z}\|_{\mathrm{b}\infty} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, \|Q\boldsymbol{z}\|_\diamond \leq 1 \right\}$$
$$= \max_{\boldsymbol{z},\boldsymbol{v}} \left\{ \|\boldsymbol{z}\|_{\mathrm{b}\infty} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, Q\boldsymbol{z} = \boldsymbol{v}, \|\boldsymbol{v}\|_\diamond \leq 1 \right\}$$
$$= \max_{\boldsymbol{z},\boldsymbol{v},\boldsymbol{u}} \left\{ \boldsymbol{u}^T \boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, Q\boldsymbol{z} = \boldsymbol{v}, \|\boldsymbol{v}\|_\diamond \leq 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1 \right\}$$
$$= \max_{\boldsymbol{z},\boldsymbol{v},\boldsymbol{u}} \left\{ \boldsymbol{u}^T (\boldsymbol{z} - P^T(Q\boldsymbol{z} - \boldsymbol{v})) : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, Q\boldsymbol{z} = \boldsymbol{v}, \|\boldsymbol{v}\|_\diamond \leq 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1 \right\}$$
$$\leq \max_{\boldsymbol{z},\boldsymbol{v},\boldsymbol{u}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)\boldsymbol{z} + \boldsymbol{u}^T P^T \boldsymbol{v} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, \|\boldsymbol{v}\|_\diamond \leq 1, \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1 \right\} (113)$$

where for the last inequality we have dropped the constraint $Q\boldsymbol{z} = \boldsymbol{v}$. Similar to (85), we bound from above the first term in the objective function of (113)

$$\max_{\boldsymbol{z},\boldsymbol{u}} \left\{ \boldsymbol{u}^T (\mathrm{I}_{np} - P^T Q)\boldsymbol{z} : \|\boldsymbol{z}\|_{\mathrm{b}1} \leq s\eta, \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1 \right\}$$
$$\leq \max_{i,j} s\eta \|(\mathrm{I}_{np} - P^T Q)_{i,j}\|_2$$
$$= \max_{i,j} s\eta \|\delta_{ij} \mathrm{I}_n - P_i^T Q_j\|_2. \qquad (114)$$

Here $P_i$ (resp. $Q_j$) is the submatrix of $P$ (resp. $Q$) formed by the $((i-1)n+1)$th to $in$th columns (resp. $((j-1)n+1)$th to $jn$th columns). For the second term $\boldsymbol{u}^T P^T \boldsymbol{v}$ in the objective function of (113), the definition of $\|\cdot\|_\diamond^*$, the dual norm of $\|\cdot\|_\diamond$, leads to

$$\max_{\boldsymbol{u},\boldsymbol{v}} \left\{ \boldsymbol{u}^T P^T \boldsymbol{v} : \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1, \|\boldsymbol{v}\|_\diamond \leq 1 \right\}$$
$$\leq \max_{\boldsymbol{u}} \left\{ \|P\boldsymbol{u}\|_\diamond^* : \|\boldsymbol{u}\|_{\mathrm{b}1} \leq 1 \right\}. \qquad (115)$$

Since $\|P\boldsymbol{u}\|_\diamond^*$ is a convex function of $\boldsymbol{u}$ and the extremal points of the unit ball $\{\boldsymbol{u} : \|\boldsymbol{u}\|_{\mathrm{b1}} \leq 1\}$ are $\{\mathbf{e}_i^p \otimes \boldsymbol{v} : \|\boldsymbol{v}\|_2 \leq 1, \boldsymbol{v} \in \mathbb{R}^n\}$, the above inequality is further bounded from above as

$$\max_{\boldsymbol{u}} \{\|P\boldsymbol{u}\|_\diamond^* : \|\boldsymbol{u}\|_{\mathrm{b1}} \leq 1\}$$
$$\leq \max_i \max_{\boldsymbol{v}} \{\|P_i\boldsymbol{v}\|_\diamond^* : \|\boldsymbol{v}\|_2 \leq 1\} = \max_i \|P_i\|_{2,*}. \qquad (116)$$

Here $\|P_i\|_{2,*}$ 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 $\max_{\boldsymbol{z}} \{\|\boldsymbol{z}\|_{\mathrm{b\infty}} : \|\boldsymbol{z}\|_{\mathrm{b1}} \leq s\eta, \|Q\boldsymbol{z}\|_\diamond \leq 1\}$ is obtained by minimizing with respect to $P$:

$$\max_{\boldsymbol{z}} \{\|\boldsymbol{z}\|_{\mathrm{b\infty}} : \|\boldsymbol{z}\|_{\mathrm{b1}} \leq s\eta, \|Q\boldsymbol{z}\|_\diamond \leq 1\}$$
$$\leq \min_P \max_{i,j} s\eta\|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2 + \|P_i\|_{2,*}$$
$$= \max_i \min_{P_i} \max_j s\eta\|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2 + \|P_i\|_{2,*}. \qquad (117)$$

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

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

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

$$\|P_i\|_{2,*} = \|P_i\|_{2,\mathrm{b1}} = \max_{\boldsymbol{v}:\|\boldsymbol{v}\|_2 \leq 1} \|P_i\boldsymbol{v}\|_{\mathrm{b1}}$$

$$= \max_{\boldsymbol{v}:\|\boldsymbol{v}\|_2 \leq 1} \sum_{l=1}^p \|P_i^l\boldsymbol{v}\|_2 \leq \sum_{l=1}^p \|P_i^l\|_2. \qquad (119)$$

Here $P_i^l$ is the submatrix of $P$ formed by the $((i-1)n+1)$th to $in$th columns and the $((l-1)n+1)$th to $ln$th rows. Proposition 4 then follows from (117), (118), and (119). $\qquad\square$

## 8.9 Proof of Proposition 5

*Proof* 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 $P_i = 0$:

$$g_{s,i} = \min_{P_i}\{s\eta(\max_j \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2) + \|P_i\|_2 : \|P_i\|_2 \leq s\eta\}$$
$$= -\max_{P_i}\{-s\eta(\max_j \|\delta_{ij}\mathrm{I}_n - P_i^T Q_j\|_2) - \|P_i\|_2 : \|P_i\|_2 \leq s\eta\}.(120)$$

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

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

is compact-valued and continuous (both upper and lower hemicontinuous), according to Berge's Maximum Theorem [1] the optimal value function $g_{s,i}(\eta)$ is continuous. The continuity of $g_s(\eta)$ follows from the fact that finite maximization preserves the continuity.

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

$$
\begin{aligned}
g_{s,i}(\eta_1) &\leq s\eta_1 \left( \max_j \| \delta_{ij} I_n - P^{*T}_{i2} Q_j \|_2 \right) + \| P^*_{i2} \|_2 \\
&< s\eta_2 \left( \max_j \| \delta_{ij} I_n - P^{*T}_{i2} Q_j \|_2 \right) + \| P^*_{i2} \|_2 \\
&= g_{s,i}(\eta_2).
\end{aligned}
\tag{122}
$$

The strict increasing of $g_s(\eta)$ then follows immediately.

3. 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 $\eta$.

4. Next we show that when $\eta > 0$ is sufficiently small $g_s(\eta) \geq s\eta$. For any $i$, we have the following,

$$
\begin{aligned}
g_{s,i}(\eta) &= \min_{P_i} s\eta \left( \max_j \| \delta_{ij} I_n - P^T_i Q_j \|_2 \right) + \| P_i \|_2 \\
&\geq \min_{P_i} s\eta \left( 1 - \| P^T_i Q_i \|_2 \right) + \| P_i \|_2 \\
&\geq \min_{P_i} s\eta \left( 1 - \| P_i \|_2 \| Q_i \|_2 \right) + \| P_i \|_2 \\
&= s\eta + \min_{P_i} \| P_i \|_2 \left( 1 - s\eta \| Q_i \|_2 \right) \\
&\geq s\eta > \eta,
\end{aligned}
\tag{123}
$$

where the minimum of the last optimization is achieved at $P_i = 0$ when $\eta < 1/(s\|Q_i\|_2)$. Clearly, $g_s(\eta) = \max_i g_{s,i}(\eta) \geq s\eta > \eta$ for such $\eta$. Recall that

$$
\frac{1}{s_*} = \max_i \min_{P_i} \max_j \| \delta_{ij} I_n - P^T_i Q_j \|_2.
\tag{124}
$$

Suppose $P^*_i$ is the optimal solution for each $\min_{P_i} \max_j \| \delta_{ij} I_n - P^T_i Q_j \|_2$. For each $i$, we then have

$$
\frac{1}{s_*} \geq \max_j \| \delta_{ij} I_n - P^{*T}_i Q_j \|_2,
\tag{125}
$$

which implies

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

As a consequence, we obtain

$$
g_s(\eta) = \max_i g_{s,i}(\eta) \leq \frac{s}{s_*} \eta + \max_i \| P^*_i \|_2.
\tag{127}
$$

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

$$g_s(\eta) \le \rho\eta. \tag{128}$$

5. We first show the existence and uniqueness of the positive fixed points for $g_{s,i}(\eta)$. Properties 1) and 4) imply that $g_{s,i}(\eta)$ has at least one positive fixed point. (Interestingly, 2) and 4) also imply the existence of a positive fixed point, see [22].) To prove uniqueness, suppose there are two fixed points $0 < \eta_1^* < \eta_2^*$. Pick $\eta_0$ small enough such that $g_{s,i}(\eta_0) > \eta_0 > 0$ and $\eta_0 < \eta_1^*$. Then $\eta_1^* = \lambda\eta_0 + (1-\lambda)\eta_2^*$ for some $\lambda \in (0,1)$, which implies that $g_{s,i}(\eta_1^*) \ge \lambda g_{s,i}(\eta_0) + (1-\lambda)g_{s,i}(\eta_2^*) > \lambda\eta_0 + (1-\lambda)\eta_2^* = \eta_1^*$ due to the concavity, contradicting $\eta_1^* = g_{s,i}(\eta_1^*)$.

The set of positive fixed points for $g_s(\eta)$, $\{\eta \in (0,\infty) : \eta = g_s(\eta) = \max_i g_{s,i}(\eta)\}$, is a subset of $\bigcup_{i=1}^p \{\eta \in (0,\infty) : \eta = g_{s,i}(\eta)\} = \{\eta_i^*\}_{i=1}^p$. We argue that

$$\eta^* = \max_i \eta_i^* \tag{129}$$

is the unique positive fixed point for $g_s(\eta)$.

We proceed to show that $\eta^*$ is a fixed point of $g_s(\eta)$. Suppose $\eta^*$ is a fixed point of $g_{s,i_0}(\eta)$. Then it suffices to show that $g_s(\eta^*) = \max_i g_{s,i}(\eta^*) = g_{s,i_0}(\eta^*)$. If this is not the case, there exists $i_1 \ne i_0$ such that $g_{s,i_1}(\eta^*) > g_{s,i_0}(\eta^*) = \eta^*$. The continuity of $g_{s,i_1}(\eta)$ and the property 4) imply that there exists $\eta > \eta^*$ with $g_{s,i_1}(\eta) = \eta$, contradicting the definition of $\eta^*$.

To show the uniqueness, suppose $\eta_1^*$ is fixed point of $g_{s,i_1}(\eta)$ satisfying $\eta_1^* < \eta^*$. Then, we must have $g_{s,i_0}(\eta_1^*) > g_{s,i_1}(\eta_1^*)$, because otherwise the continuity implies the existence of another fixed point of $g_{s,i_0}(\eta)$. As a consequence, $g_s(\eta_1^*) > g_{s,i_1}(\eta_1^*) = \eta_1^*$ and $\eta_1^*$ is not a fixed point of $g_s(\eta)$.

6. This property simply follows from the continuity, the uniqueness, and property 4).

7. We use contradiction to show the existence of $\rho_1(\epsilon)$ in 7). In view of 4), we need only to show the existence of such a $\rho_1(\epsilon)$ that works for $\eta_L \le \eta \le (1-\epsilon)\eta^*$ where $\eta_L = \sup\{\eta : g_s(\xi) > s\xi, \forall 0 < \xi \le \eta\}$. Supposing otherwise, we then construct sequences $\{\eta^{(k)}\}_{k=1}^\infty \subset [\eta_L, (1-\epsilon)\eta^*]$ and $\{\rho_1^{(k)}\}_{k=1}^\infty \subset (1,\infty)$ with

$$\lim_{k\to\infty} \rho_1^{(k)} = 1, \text{ and}$$
$$g_s(\eta^{(k)}) \le \rho^{(k)}\eta^{(k)}. \tag{130}$$

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

$$g_s(\eta_{\lim}) = \lim_{l\to\infty} g_s(\eta^{(k_l)}) \le \lim_{l\to\infty} \rho_1^{(k_l)}\eta^{(k_l)} = \eta_{\lim}. \tag{131}$$

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

$$g_s(\eta_c) = \eta_c, \tag{132}$$

contradicting the uniqueness of the fixed point for $g_s(\eta)$. The existence of $\rho_2(\epsilon)$ can be proved in a similar manner.

$\square$

## References

1. Berge, C.: Topological Spaces. Dover Publications, Mineola, NY (reprinted 1997 in paperback)
2. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. Compte Rendus de l'Academie des Sciences, Paris, Serie I **346**, 589–592 (2008)
3. Candès, E.J., Plan, Y.: Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. IEEE Trans. Inf. Theory **57**(4), 2342–2359 (2011)
4. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
5. d'Aspremont, A., El Ghaoui, L.: Testing the nullspace property using semidefinite programming. Mathematical Programming **127**(1), 123–144 (2011)
6. d'Aspremont, A., Ghaoui, L.E., Jordan, M., Lanckriet, G.R.G.: A direct formulation for sparse PCA using semidefinite programming. SIAM Review **49**(3), 434–448 (2007)
7. Eldar, Y.C., Mishali, M.: Robust recovery of signals from a structured union of subspaces. IEEE Trans. Inf. Theory **55**(11), 5302 –5316 (2009)
8. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming (web page and software) (2009). URL `http://stanford.edu/~boyd/cvx`
9. Juditsky, A., Nemirovski, A.: On verifiable sufficient conditions for sparse signal recovery via $\ell_1$ minimization. Mathematical Programming **127**(1), 57–88 (2011)
10. Ledoux, M., Talagrand, M.: Probability in Banach spaces: Isoperimetry and processes. Springer-Verlag, New York (1991)
11. Liu, H., Zhang, J., Jiang, X., Liu, J.: The group dantzig selector. In: Inter. Conf. Artificial Intelligence and Statistics (AISTATS 2010). Sardinia, Italy (2010)
12. Malioutov, D., Cetin, M., Willsky, A.S.: A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005)
13. Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Reconstruction and subgaussian operators in asymptotic geometric analysis. Geometric And Functional Analysis pp. 1248–1282 (2007)
14. Mishali, M., Eldar, Y.C.: Blind multi-band signal reconstruction: Compressed sensing for analog signals. IEEE Trans. Signal Process. **57**(3), 993–1009 (2009)
15. Nesterov, Y.: Smoothing technique and its applications in semidefinite optimization. Math. Program. **110**, 245–259 (2007)
16. Parvaresh, F., Vikalo, H., Misra, S., Hassibi, B.: Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays. IEEE J. Sel. Topics Signal Processing **2**(3), 275–285 (2008)
17. Sen, S., Nehorai, A.: Sparsity-based multi-target tracking using OFDM radar. IEEE Trans. Signal Process. **59**, 1902–1906 (2011)
18. Sen, S., Tang, G., Nehorai, A.: Multiobjective optimization of OFDM radar waveform for target detection. IEEE Trans. Signal Process. **59**(2), 639–652 (2011)
19. Stojnic, M., Parvaresh, F., Hassibi, B.: On the reconstruction of block-sparse signals with an optimal number of measurements. IEEE Trans. Signal Process. **57**(8), 3075–3085 (2009)
20. Tang, G., Nehorai, A.: Performance analysis of sparse recovery based on constrained minimal singular values. to appear in IEEE Trans. Signal Processing

21. Tang, G., Nehorai, A.: Verifiable and computable $\ell_\infty$ performance evaluation of $\ell_1$ sparse signal recovery. submitted to SIAM J. Scientific Computing
22. Tarski, A.: A lattice-theoretical fix point theorem and its applications. Pacific Journal of Mathematics **5**, 285–309 (1955)
23. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: Y. Eldar, G. Kutyniok (eds.) Compressed Sensing: Theory and Applications (to appear)
24. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **68**(1), 49–67 (2006)