

368 Contents

369	6 Appendix A: Notations and definitions	12
370	6.1 Mixed $\ell_{p,q}$ norm of a matrix	12
371	6.2 Mixed $\ell_{p,q}$ norm over block partition of a vector	12
372	7 Appendix B: Preliminaries	13
373	7.1 Block Sparsity	13
374	7.2 Null Space Property (NSP) and Block-NSP (BNSP)	13
375	7.3 Restricted Isometry Property (RIP) and Block-RIP (BRIP)	13
376	7.4 Noisy Recovery bounds based on RIP constants	14
377	7.5 Sample Complexity for i.i.d. Gaussian or Bernoulli Random Matrices	14
378	8 Appendix C: Proofs of JOBS theorems	14
379	8.1 Proof of Theorem 2: correctness of JOBS	14
380	8.2 Proof of Theorem 4: JOBS performance bound of for exactly s -sparse signals . .	15
381	8.3 Proof of JOBS performance bound for general sparse signals	17
382	9 Appendix D: Proofs of Bagging theorems	19
383	9.1 Proof of Bagging Performance bound for exactly s -sparse signals	19
384	9.2 Proof of Bagging performance bound of bagging for approximately sparse signals .	21
385	10 Appendix E: Theory for NSP and RIP	22
386	10.1 Proof of the reverse direction for noiseless recovery	22
387	10.2 JOBS matrix satisfies BNSP implies that each block matrix satisfies NSP	23
388	10.3 Proof of Proposition 3	23
389	10.4 Proof of Lemma 17	24
390	11 Appendix F: Pseudo-code of JOBS implementation via ADMM	25
391	12 Appendix G: Distribution of the unique number of elements for bootstrapping	25
392	12.0.1 Unique number of Bootstrap Samples with Finite Sample m	26
393	12.0.2 Asymptotic unique ratios of bootstrap samples	27
394	12.0.3 Finite number of measurements m cases are empirically close to the asymp-	
395	totic case	27
396	12.1 The sub-sampling Variation: Sub-JOBS	27
397	13 Appendix H: JOBS, a two step relaxation of ℓ_1 minimization	27

398 6 Appendix A: Notations and definitions

The full list of notations that are used in this paper is in Table 4.

Table 4: Notation of Main Variables

m	total number of measurements
n	signal dimension
s	sparsity level
L	size of each bootstrap sample
L/m	bootstrap sampling ratio
K	number of bootstrap samples / the number of estimates
\mathbf{A}	the original sensing matrix of size $m \times n$
\mathbf{y}	the original measurements vector of size $m \times 1$
\mathcal{I}	a multi-set or set
\mathcal{I}_j	the j -th Bootstrap sample, $j = 1, 2, \dots, K$, length of $\mathcal{I}_j = L$
$(\cdot)[\mathcal{I}]$	takes rows supported on \mathcal{I} and throws away elements in \mathcal{I}^c
$\mathbf{A}[\mathcal{I}_j]$	bootstrapped sampling matrix for bootstrap sample \mathcal{I}_j
$\mathbf{y}[\mathcal{I}_j]$	measurement vector corresponds to bootstrap sample \mathcal{I}_j
\mathbf{x}_j	the j -th column of matrix \mathbf{X} ; a feasible solution corresponds to $(\mathbf{A}[\mathcal{I}_j], \mathbf{y}[\mathcal{I}_j])$
$\hat{\mathbf{x}}_j$	the optimal solution corresponds to $(\mathbf{A}[\mathcal{I}_j], \mathbf{y}[\mathcal{I}_j])$
$(\cdot)[i]$	the i -th row of a matrix/ vector.
$\mathbf{x}[i]$	the i -th row of matrix \mathbf{X}
$\ \mathbf{X}\ _{p,q}$	takes ℓ_q norms on rows of \mathbf{X} ; stacks those as a vector and then computes ℓ_p norm. The precise form is in (8).
$\ \mathbf{X}\ _{1,2}$	row sparsity norm
$\ \mathbf{X}\ _{1,1}$	the ℓ_1 norm on vectorized \mathbf{X}

399

400 6.1 Mixed $\ell_{p,q}$ norm of a matrix

401 The mixed $\ell_{p,q}$ norm on matrix \mathbf{X} is defined as:

$$\begin{aligned} \|\mathbf{X}\|_{p,q} &= \left(\sum_{i=1}^n \|\mathbf{x}[i]^T\|_q^p \right)^{1/p} \\ &= \|(\|\mathbf{x}[1]^T\|_q, \|\mathbf{x}[2]^T\|_q, \dots, \|\mathbf{x}[n]^T\|_q)^T\|_p, \end{aligned} \quad (8)$$

402 where $\mathbf{x}[i]$ denotes the i -th row of matrix \mathbf{X} . Intuitively, the mixed $\ell_{p,q}$ norm essentially takes ℓ_q
403 norms on rows of \mathbf{X} first; then stacks those as a vector and then computes its ℓ_p norm. Note when
404 $p = q$, the $\ell_{p,p}$ norm of $\|\mathbf{X}\|$ is simply the ℓ_p vector norm of the vectorized \mathbf{X} . The row sparsity
405 penalty that we employed $\ell_{1,2}$ norm in JOBS is essentially a special case of (8) taking $p = 1, q = 2$.

406 6.2 Mixed $\ell_{p,q}$ norm over block partition of a vector

407 Similarly to the $\ell_{p,q}$ norm on matrix in (8), we introduce a more general form: the mixed $\ell_{p,q}$
408 norm over a block partition of a vector. The definition for $\ell_{p,q}$ norm over block partition $\mathcal{B} =$
409 $\{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_b\}$ for a vector $\|\mathbf{x}\|_{p,q|\mathcal{B}}$:

$$\begin{aligned} \|\mathbf{x}\|_{p,q|\mathcal{B}} &= \left(\sum_{i=1}^b \|\mathbf{x}[\mathcal{B}_i]^T\|_q^p \right)^{1/p} \\ &= \|(\|\mathbf{x}[\mathcal{B}_1]^T\|_q, \dots, \|\mathbf{x}[\mathcal{B}_b]^T\|_q)\|_p. \end{aligned} \quad (9)$$

It is not difficult to see that the $\ell_{p,q}$ norm of a matrix is a special case of $\ell_{p,q}$ norm over block of the vectorized version of that matrix. In fact, the mixed $\ell_{1,2}$ norm on matrix \mathbf{X} can also be expressed as a mixed $\ell_{1,2|\mathcal{B}}$ norm on the vectorized \mathbf{X} given \mathcal{B} , where the block partition is row-wise.

7 Appendix B: Preliminaries

We summarize the theoretical results that are needed for understanding and analyzing our algorithm mathematically. We offer a quick review of several concepts including block sparsity, Null Space Property (NSP) [1], Restricted Isometry Property (RIP) [2] for classical sparse signal recovery as well as Block Null Space Property (BNSP) [21], Block Restricted Isometry Property (BRIP) [17] for block sparse signal recovery.

7.1 Block Sparsity

Since row sparsity is a special case of block sparsity (or more precisely, the non-overlapping group sparsity) [17], we therefore can employ the tools from block sparsity to analyze our problem. Block sparsity is a generalization of the standard ℓ_1 sparsity. To start, we recall its definition.

Definition 8 (Block Sparsity, from [17]) $\mathbf{x} \in \mathbb{R}^n$ is s -block sparse with respect to a partition $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_b\}$ of $\{1, 2, \dots, n\}$ if for $\mathbf{x} = (\mathbf{x}[\mathcal{B}_1], \mathbf{x}[\mathcal{B}_2], \dots, \mathbf{x}[\mathcal{B}_b])$, the block sparsity level is $\|\mathbf{x}\|_{0,2|\mathcal{B}} := \sum_{i=1}^b 1\{\|\mathbf{x}[\mathcal{B}_i]\|_2 > 0\} \leq s$ and the relaxation $\ell_{1,2}$ norm is $\|\mathbf{x}\|_{1,2|\mathcal{B}} := \sum_{i=1}^b \|\mathbf{x}[\mathcal{B}_i]\|_2$.

The block sparsity level $\|\mathbf{x}\|_{0,2|\mathcal{B}}$ counts the number of non-zero blocks of the given a block partition \mathcal{B} . The $\ell_{1,2}$ norm $\|\mathbf{x}\|_{1,2|\mathcal{B}} := \sum_{i=1}^b \|\mathbf{x}[\mathcal{B}_i]\|_2$ is one of its convex relaxations. For the same sparse vector \mathbf{x} , the block sparsity level is in general smaller than the sparsity level given a non-overlapping block partition.

The $\ell_{1,2}$ minimization is a special case of block sparse minimization, with each element in the block partition containing all indices of a row. The results of block sparsity such as BNSP, BRIP can be useful tools to analyze our algorithm.

7.2 Null Space Property (NSP) and Block-NSP (BNSP)

The NSP for standard sparse recovery and block sparse signal recovery are summarized below. BNSP is obtained from a more general result of BNSP of $\ell_{p,2}$ block norm stated in (9) from [21] taking $p = 1$.

Theorem 9 (NSP, from [1]) Every s -sparse signal $\mathbf{x} \in \mathbb{R}^n$ is a unique solution to $\mathbf{P}_1 : \min \|\mathbf{x}\|_1$ s.t. $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if \mathbf{A} satisfies NSP of order s : for any set $\mathcal{S} \subset \{1, 2, \dots, n\}$, $\text{card}(\mathcal{S}) \leq s$,

$$\|\mathbf{v}[\mathcal{S}]\|_1 < \|\mathbf{v}[\mathcal{S}^c]\|_1,$$

for all $\mathbf{v} \in \text{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, where $\mathbf{v}[\mathcal{S}]$ denotes the vector equals to \mathbf{v} on a index set \mathcal{S} and zero elsewhere.

Definition 10 (BNSP, from [21]) Every s -block sparse signal \mathbf{x} with respect to block assignment \mathcal{B} , is a unique solution to $\min \|\mathbf{x}\|_{1,2|\mathcal{B}}$ s.t. $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if matrix \mathbf{A} satisfies block null space property over \mathcal{B} of order s : for any set $\mathcal{S} \subset \{1, 2, \dots, n\}$ with $\text{card}(\mathcal{S}) \leq s$,

$$\|\mathbf{v}[\mathcal{S}]\|_{1,2|\mathcal{B}} < \|\mathbf{v}[\mathcal{S}^c]\|_{1,2|\mathcal{B}},$$

for all $\mathbf{v} \in \text{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, where $\mathbf{v}[\mathcal{S}]$ denotes the vector equal to \mathbf{v} on a block index set \mathcal{S} and zero elsewhere.

7.3 Restricted Isometry Property (RIP) and Block-RIP (BRIP)

Although NSP directly characterizes the ability of success for sparse recovery, verifying the BNSP condition is computationally intractable and it is also not suitable for quantifying performance in noisy cases since it is a binary (True or False) metric instead of a continuous one. Restricted Isometry Properties: RIP [2] and BRIP [17] are introduced for those purposes.

453 **Definition 11 (RIP, from [2])** A matrix \mathbf{A} with ℓ_2 -normalized columns satisfies RIP of order s if
 454 there exists a constant $\delta_s(\mathbf{A}) \in [0, 1)$ such that for every s -sparse $\mathbf{v} \in \mathbb{R}^n$,

$$(1 - \delta_s(\mathbf{A}))\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1 + \delta_s(\mathbf{A}))\|\mathbf{v}\|_2^2. \quad (10)$$

455 **Definition 12 (BRIP, from [17])** A matrix \mathbf{A} with ℓ_2 -normalized columns satisfies Block RIP with
 456 respect to block partition \mathcal{B} of order s if there exists a constant $\delta_{s|\mathcal{B}}(\mathbf{A}) \in [0, 1)$ such that for every
 457 s -block sparse $\mathbf{v} \in \mathbb{R}^n$ over \mathcal{B} ,

$$(1 - \delta_{s|\mathcal{B}}(\mathbf{A}))\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1 + \delta_{s|\mathcal{B}}(\mathbf{A}))\|\mathbf{v}\|_2^2. \quad (11)$$

458 If we take the location of each entry as one block, the block sparsity RIP reduces to the standard RIP
 459 condition. Therefore, BRIP is a generalization of RIP.

460 7.4 Noisy Recovery bounds based on RIP constants

461 It is well-known that certain RIP conditions imply NSP conditions for both classical sparse recovery
 462 and block sparse recovery. More specifically, if the RIP constant in the order $2s$ is strictly less than
 463 $\sqrt{2} - 1$, then it implies that NSP is satisfied in the order of s . This applies to sparse recovery [2] and
 464 block sparse recovery [17].

465 Stated below are the error bound for conventional sparse recovery based on ℓ_1 minimization and the
 466 RIP constant as well as for block sparse recovery based on BRIP constant.

467 **Theorem 13 (Sparse recovery error bound, from [2])** Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{z}$, $\|\mathbf{z}\|_2 \leq \epsilon$; \mathbf{x}_0 is
 468 s -sparse and minimizes $\|\mathbf{x} - \mathbf{x}^*\|_2$ over all s -sparse signals, and the vector \mathbf{e} represents the
 469 s -sparse approximation error vector $\mathbf{e} = \mathbf{x}^* - \mathbf{x}_0$. If $\delta_{2s}(\mathbf{A}) \leq \delta < \sqrt{2} - 1$ and \mathbf{x}^{ℓ_1} is the solution
 470 of ℓ_1 minimization, then

$$\|\mathbf{x}^{\ell_1} - \mathbf{x}^*\|_2 \leq C_0(\delta)s^{-1/2}\|\mathbf{e}\|_1 + C_1(\delta)\epsilon,$$

471 where $C_0(\cdot), C_1(\cdot)$ are certain constants, depending on the RIP constant $\delta_{2s}(\mathbf{A})$. These two constants
 472 are in the form of non-decreasing functions of δ : $C_0(\delta) = \frac{2(1-(1-\sqrt{2})\delta)}{1-(1+\sqrt{2})\delta}$ and $C_1(\delta) = \frac{4\sqrt{1+\delta}}{1-(1+\sqrt{2})\delta}$.

473 **Theorem 14 (Block sparse recovery error bound, from [17])** Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{z}$, $\|\mathbf{z}\|_2 \leq \epsilon$; $\mathbf{x}_{0|\mathcal{B}}$
 474 is s -block sparse and minimizes $\|\mathbf{x} - \mathbf{x}^*\|_2$ over all s -block sparse signals, and the vector $\mathbf{e}_{\mathcal{B}}$
 475 represents the s -sparse approximation error vector $\mathbf{e}_{\mathcal{B}} = \mathbf{x}^* - \mathbf{x}_{0|\mathcal{B}}$. If $\delta_{2s|\mathcal{B}}(\mathbf{A}) \leq \delta < \sqrt{2} - 1$,
 476 $\mathbf{x}^{\ell_{1,2|\mathcal{B}}}$ is the solution of block sparse minimization, then

$$\|\mathbf{x}^{\ell_{1,2|\mathcal{B}}} - \mathbf{x}^*\|_2 \leq C_0(\delta)s^{-1/2}\|\mathbf{e}_{\mathcal{B}}\|_{1,2|\mathcal{B}} + C_1(\delta)\epsilon,$$

477 where $C_0(\cdot), C_1(\cdot)$ are the same non-decreasing functions of δ as in Theorem 13.

478 7.5 Sample Complexity for i.i.d. Gaussian or Bernoulli Random Matrices

479 With \mathbf{A} being a random matrix in which entries are identically and independently distributed (i.i.d.),
 480 previous work in [18] builds a relationship between the sample complexity for random matrices to a
 481 desired RIP constant as a direct implication from Johnson-Lindenstrauss lemma as stated below.

482 **Theorem 15 (Sample Complexity, from [18])** Let entries of $\mathbf{A} \in \mathbb{R}^{m \times n}$ from Gaussian
 483 $\mathcal{N}(0, 1/m)$ or Bernoulli $1/\sqrt{m}$ Bern(0.5). Let $\xi, \delta \in (0, 1)$ and assume $m \geq \beta\delta^{-2}(s \ln(n/s) +$
 484 $\ln(\xi^{-1}))$ for a universal constant $\beta > 0$, then $\mathbb{P}(\delta_s(\mathbf{A}) \leq \delta) \geq 1 - \xi$.

485 By rearranging the terms in this theorem, the sample complexity can be derived: when m is sufficiently
 486 large, which is in the order of $\mathcal{O}(2s \ln(n/2s))$, there is a high probability that the RIP constant of
 487 order $2s$ is sufficiently small.

488 8 Appendix C: Proofs of JOBS theorems

489 8.1 Proof of Theorem 2: correctness of JOBS

490 The first part of Theorem 2 can be directly shown from the BNSP for block sparse minimization
 491 problems as in [17]. We only need to show the procedure to prove the latter part. If BNSP of order s

is satisfied for $\{\mathbf{A}[\mathcal{I}_1], \mathbf{A}[\mathcal{I}_2], \dots, \mathbf{A}[\mathcal{I}_K]\}$, then each bootstrap matrix $\mathbf{A}[\mathcal{I}_j]$ satisfies the Null Space Property (NSP) of order s , which is proven in Appendix 10.2. Consequently, for all $j = 1, 2, \dots, K$, \mathbf{x}^* also turns out to be the optimal solution to all sub-problems: $\mathbf{x}^* = \arg \min_{\mathbf{x}_j} \|\mathbf{x}_j\|_1$ s.t. $\mathbf{y}[\mathcal{I}_j] = \mathbf{A}[\mathcal{I}_j]\mathbf{x}_j$.

For \mathbf{X} to be a feasible solution, consider its $\ell_{1,2}$ norm, we have:

$$\|\mathbf{X}\|_{1,2} = \sum_{i=1}^n \left(\sum_{j=1}^K (x_{ij}^2) \right)^{1/2} = \sqrt{K} \sum_{i=1}^n \left(\frac{1}{K} \sum_{j=1}^K (x_{ij}^2) \right)^{1/2}.$$

By concavity of the square root, we have

$$\begin{aligned} \|\mathbf{X}\|_{1,2} &\geq \sqrt{K} \sum_{i=1}^n \frac{1}{K} \sum_{j=1}^K \sqrt{x_{ij}^2} = \sqrt{K} \frac{1}{K} \sum_{j=1}^K \sum_{i=1}^n |x_{ij}| \\ &\geq \sqrt{K} \frac{1}{K} \sum_{j=1}^K \min_{\mathbf{x}_j: \mathbf{x}_{1,j}, \dots, \mathbf{x}_{n,j} \in \mathbf{A}[\mathcal{I}_j] \mathbf{x}_j = \mathbf{y}[\mathcal{I}_j]} \sum_{i=1}^n |x_{ij}| \\ &= \sqrt{K} \frac{1}{K} \sum_{j=1}^K \min_{\mathbf{x}_j: \mathbf{A}[\mathcal{I}_j] \mathbf{x}_j = \mathbf{y}[\mathcal{I}_j]} \|\mathbf{x}_j\|_1 \\ &= \sqrt{K} \|\mathbf{x}^*\|_1. \end{aligned}$$

Since $\mathbf{X}^* = (\mathbf{x}^*, \mathbf{x}^*, \dots, \mathbf{x}^*)$ is a feasible solution and $\|\mathbf{X}^*\|_{1,2} = \|(\mathbf{x}^*, \mathbf{x}^*, \dots, \mathbf{x}^*)\|_{1,2} = \sqrt{K} \|\mathbf{x}^*\|_1$, it achieves the lower bound. By the uniqueness part of the theorem, we can concluded that \mathbf{X}^* is the unique solution. Since the JOBS solution takes the average over columns of multiple estimates, we can easily deduce that JOBS returns the correct answer.

8.2 Proof of Theorem 4: JOBS performance bound of for exactly s -sparse signals

If the true solution is exactly s -sparse, the sparse approximation error is zero. Then the noise level of performance only relates to measurements noise. For ℓ_1 minimization, \mathbf{z} is the noise vector and we use matrix $\mathbf{Z} = (\mathbf{z}[\mathcal{I}_1], \mathbf{z}[\mathcal{I}_2], \dots, \mathbf{z}[\mathcal{I}_K])$ to denote the noise matrix in JOBS. We bound the distance of $\|\mathbf{Z}\|_{2,2}$ to its expected value using Hoeffding's inequalities stated in [22].

Theorem 16 (Hoeffding's Inequalities) *Let X_1, \dots, X_n be independent bounded random variables such that X_i falls in the interval $[a_i, b_i]$ with probability one. Denote their sum by $S_n = \sum_{i=1}^n X_i$. Then for any $\zeta > 0$, we have:*

$$\mathbb{P}\left\{S_n - \mathbb{E}S_n \geq \zeta\right\} \leq \exp \frac{-2\zeta^2}{\sum_{i=1}^n (b_i - a_i)^2} \quad \text{and} \quad (12)$$

$$\mathbb{P}\left\{S_n - \mathbb{E}S_n \leq -\zeta\right\} \leq \exp \frac{-2\zeta^2}{\sum_{i=1}^n (b_i - a_i)^2}. \quad (13)$$

Here, the entire noise vector is $\mathbf{z} = \mathbf{A}\mathbf{x} - \mathbf{y} = (z[1], z[2], \dots, z[m])^T$, $\|\mathbf{z}\|_\infty = \max_{i=1,2,\dots,m} |z[i]| < \infty$. We consider the matrix $\mathbf{Z} \circ \mathbf{Z} = (\xi_{ji})$, where \circ is the entry-wise product. The quantity that we are interested in $\|\mathbf{Z}\|_{2,2}$ is the sum of all entries in $\mathbf{Z} \circ \mathbf{Z}$. Each element in this matrix $\mathbf{Z} \circ \mathbf{Z}$ is drawn i.i.d from the squares of entries in \mathbf{z} : $\{z[1], z[2], \dots, z[m]\}$ with equal probability. Let Ξ be the underlining random variable and Ξ obeys a discrete uniform distribution:

$$\mathbb{P}(\Xi = z^2[i]) = \frac{1}{m}, i = 1, 2, \dots, m. \quad (14)$$

The lower and upper bound of Ξ is then

$$0 \leq \min_i z^2[i] \leq \Xi \leq \|\mathbf{z}\|_\infty^2. \quad (15)$$

517 We use zero as lower bound for Ξ instead of the minimum value to simplify the terms. The expected
 518 power of \mathbf{Z} is

$$\mathbb{E}\|\mathbf{Z}\|_{2,2}^2 = \frac{KL}{m}\|\mathbf{z}\|_2^2. \quad (16)$$

519 Applying Hoeffding's inequality for any $\tau > 0$ leads to

$$\mathbb{P}\{\|\mathbf{Z}\|_{2,2}^2 - \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 - \tau \leq 0\} \geq 1 - \exp \frac{-2\tau^2}{KL\|\mathbf{z}\|_\infty^4}. \quad (17)$$

520 Next, let $\widehat{\mathbf{X}}$ be the solution of \mathbf{J}_{12}^λ . Theorem 14 yields

$$\mathbb{P}\{\|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_{2,2}^2 - \mathcal{C}_1^2(\delta)\|\mathbf{Z}\|_{2,2}^2 \leq 0\} = 1. \quad (18)$$

521 Let Δ denote the difference between the solution to the truth solution scaled by the \mathcal{C}_1 constant.

522 Hence, $\Delta = \frac{1}{\mathcal{C}_1(\delta)}\|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_{2,2}$ and (18) becomes

$$\mathbb{P}\{\Delta - \|\mathbf{Z}\|_{2,2} \leq 0\} = 1. \quad (19)$$

523 Since \mathbf{Z} depends on the choice of $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_K$, we derive the typical performance by studying the
 524 distance of the solution to the expected noise level of JOBS.

$$\begin{aligned} & \mathbb{P}\{\Delta^2 - \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 - \tau^2 \leq 0\} \\ &= \mathbb{P}\{\Delta^2 - \|\mathbf{Z}\|_{2,2}^2 + \|\mathbf{Z}\|_{2,2}^2 - \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 - \tau^2 \leq 0\} \\ &\geq \mathbb{P}\{\Delta^2 - \|\mathbf{Z}\|_{2,2}^2 \leq 0, \|\mathbf{Z}\|_{2,2}^2 - \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 - \tau^2 \leq 0\} \\ &\quad (\text{The first and the second parts are independent}) \\ &= \mathbb{P}\{\Delta^2 - \|\mathbf{Z}\|_{2,2}^2 \leq 0\} \mathbb{P}\{\|\mathbf{Z}\|_{2,2}^2 - \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 - \tau^2 \leq 0\} \\ &\quad (\text{using (19) and (17)}) \\ &\geq 1 - \exp \frac{-2\tau^4}{KL\|\mathbf{z}\|_\infty^4}. \end{aligned}$$

525 In summary, this procedure results in

$$\mathbb{P}\{\Delta^2 \leq \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 + \tau^2\} \geq 1 - \exp \frac{-2\tau^4}{KL\|\mathbf{z}\|_\infty^4}. \quad (20)$$

526 We can bound the squared error as follows:

$$\begin{aligned} & \mathbb{P}\{\Delta \leq (\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau\} \\ &= \mathbb{P}\{\Delta^2 \leq \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 + \tau^2 + 2\tau(\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2}\} \\ &\geq \mathbb{P}\{\Delta^2 \leq \mathbb{E}\|\mathbf{Z}\|_{2,2}^2 + \tau^2\}. \end{aligned} \quad (21)$$

527 Combining (20) and (21), we arrive at

$$\mathbb{P}\{\Delta \leq (\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau\} \geq 1 - \exp \frac{-2\tau^4}{KL\|\mathbf{z}\|_\infty^4}. \quad (22)$$

528 Since $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}^\star\|_2^2$ is convex, we can apply Jensens' inequality to establish:

$$\left\| \frac{1}{K} \sum_{j=1}^K \widehat{\mathbf{x}}_j - \mathbf{x}^\star \right\|_2^2 \leq \frac{1}{K} \sum_{j=1}^K \|\widehat{\mathbf{x}}_j - \mathbf{x}^\star\|_2^2. \quad (23)$$

529 The JOBS estimate is averaged column-wise over all estimates: $\mathbf{x}^J = \frac{1}{K} \sum_{j=1}^K \widehat{\mathbf{x}}_j$. Therefore,
 530 equation (23) is essentially

$$\mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2^2 - \frac{1}{K}\|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_{2,2}^2 \leq 0\} = 1. \quad (24)$$

531 Now, we consider the typical performance of the JOBS solution and recall that Δ denotes the
 532 difference between the solution to the truth solution scaled by the \mathcal{C}_1 constant: $\Delta = \frac{1}{\mathcal{C}_1(\delta)} \|\widehat{\mathbf{X}} -$
 533 $\mathbf{X}^\star\|_{2,2}$. We can then bound the probability of error.

$$\begin{aligned}
 & \mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 - \frac{\mathcal{C}_1(\delta)}{\sqrt{K}} ((\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau) \leq 0\} \\
 &= \mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 - \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_2 \\
 &\quad + \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_2 - \frac{\mathcal{C}_1(\delta)}{\sqrt{K}} ((\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau) \leq 0\} \\
 &\geq \mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 - \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_2 \leq 0, \\
 &\quad \Delta \leq (\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau\} \\
 &= \mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 - \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_2 \leq 0\} \\
 &\quad \mathbb{P}\{\Delta \leq (\mathbb{E}\|\mathbf{Z}\|_{2,2}^2)^{1/2} + \tau\} \quad (\text{by (24) and (22)}) \\
 &\geq 1 - \exp \frac{-2\tau^4}{KL\|\mathbf{z}\|_\infty^4}.
 \end{aligned} \tag{25}$$

534 Substituting the expected noise level derived in (16) yields

$$\begin{aligned}
 & \mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 \leq \mathcal{C}_1(\delta) \left(\sqrt{\frac{L}{m}} \|\mathbf{z}\|_2 + \frac{\tau}{\sqrt{K}} \right)\} \\
 & \geq 1 - \exp \frac{-2\tau^4}{KL\|\mathbf{z}\|_\infty^4}.
 \end{aligned}$$

535 By replacing τ/\sqrt{K} with τ , the quantity on the right hand side of the equation then becomes
 536 $1 - \exp \frac{-2K\tau^4}{L\|\mathbf{z}\|_\infty^4}$ and we have proved the theorem.

537 8.3 Proof of JOBS performance bound for general sparse signals

538 Now we consider the case that the BNSP is only satisfied for order s whereas there is no s -sparse
 539 assumption on the true solution. Therefore, the JOBS algorithm can only guarantee the correctness of
 540 the s -row-sparse part and our best hope is to be able to recover the best s -sparse approximation of
 541 the true solution. Let \mathbf{x}_0 be the best s -sparse approximation of the true solution \mathbf{x}^\star and \mathbf{e} denote the
 542 difference of the sparse approximation: $\mathbf{e} = \mathbf{x}^\star - \mathbf{x}_0$. We rewrite the measurements to include the
 543 s -sparse approximation error as part of noise: for $j = 1, 2, \dots, K$,

$$\begin{aligned}
 \mathbf{y}[\mathcal{I}_j] &= \mathbf{A}[\mathcal{I}_j] \mathbf{x}^\star + \mathbf{z}[\mathcal{I}_j] \\
 &= \mathbf{A}[\mathcal{I}_j] (\mathbf{x}_0 + (\mathbf{x}^\star - \mathbf{x}_0)) + \mathbf{z}[\mathcal{I}_j] \\
 &= \mathbf{A}[\mathcal{I}_j] \mathbf{x}_0 + \tilde{\mathbf{z}}_j,
 \end{aligned} \tag{26}$$

544 where $\tilde{\mathbf{z}}_j = \mathbf{A}[\mathcal{I}_j] (\mathbf{x}^\star - \mathbf{x}_0) + \mathbf{z}[\mathcal{I}_j] = \mathbf{A}[\mathcal{I}_j] \mathbf{e} + \mathbf{z}[\mathcal{I}_j]$.

545 To bound the distance of solution of \mathbf{J}_{12}^λ : $\widehat{\mathbf{X}}$ to the true solution \mathbf{X}^\star , we evaluate its distance to the
 546 exactly s row-sparse matrix $\mathbf{X}_0 = (\mathbf{x}_0, \mathbf{x}_0, \dots, \mathbf{x}_0)$ as an intermediate step. Since $\mathbf{e} = \mathbf{x}^\star - \mathbf{x}_0$, we
 547 have: $\mathbf{X}^\star - \mathbf{X}_0 = (\mathbf{e}, \mathbf{e}, \dots, \mathbf{e})$ and $\|\mathbf{X}_0 - \mathbf{X}^\star\|_{2,2} = \sqrt{K} \|\mathbf{e}\|_2$. Then, the distance of $\widehat{\mathbf{X}}$ to the
 548 true solution \mathbf{X}^\star can be decomposed into two components:

$$\begin{aligned}
 \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_{2,2} &= \|\widehat{\mathbf{X}} - \mathbf{X}_0 + \mathbf{X}_0 - \mathbf{X}^\star\|_{2,2} \\
 &\leq \|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{2,2} + \|\mathbf{X}_0 - \mathbf{X}^\star\|_{2,2} \\
 &= \|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{2,2} + \sqrt{K} \|\mathbf{e}\|_2.
 \end{aligned} \tag{27}$$

549 To bound the first component in (27): $\|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{2,2}$, the procedure is similar to the prove the
 550 exactly s -sparse case. We use the recovery guarantee from the row sparse recovery result in

551 Theorem 14, which gives an upper bound of this term associated with the power of the noise matrix
 552 $\tilde{\mathbf{Z}} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_K)$:

$$\begin{aligned}\|\tilde{\mathbf{Z}}\|_{2,2}^2 &= \sum_{j=1}^K \|\tilde{z}_j\|_2^2 = \sum_{j=1}^K \|\mathbf{A}[\mathcal{I}_j]\mathbf{e} + \mathbf{z}[\mathcal{I}_j]\|_2^2 \\ &= \sum_{j=1}^K \sum_{i \in \mathcal{I}_j} (\langle \mathbf{a}[i], \mathbf{e} \rangle + z[i])^2.\end{aligned}\tag{28}$$

553 Next, let $\tilde{\Xi} = (\langle \mathbf{a}[i], \mathbf{e} \rangle + z[i])^2$ with $\mathbf{a}[i], \mathbf{z}[i]$ generated uniformly from all rows of \mathbf{A} and \mathbf{z} . Since
 554 $\tilde{\Xi}$ is non-negative, $\Xi \geq 0$, the lower bound is 0. Its upper bound can be derived using the Hölders
 555 inequality:

$$\begin{aligned}\tilde{\Xi} &= (\langle \mathbf{a}[i], \mathbf{e} \rangle + z[i])^2 \leq (\|\langle \mathbf{a}[i], \mathbf{e} \rangle\|_1 + \|\mathbf{z}\|_\infty)^2 \\ &\leq (\|\mathbf{a}[i]^T\|_1 \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^2 \\ &\leq (\max_i \|\mathbf{a}[i]^T\|_1 \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^2 \\ &= (\|\mathbf{A}\|_{\infty,1} \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^2,\end{aligned}\tag{29}$$

556 where $\|\mathbf{A}\|_{\infty,1} = \max_{i=1,2,\dots,m} \|\mathbf{a}[i]^T\|_1$. Since \mathbf{A} is deterministic with all bounded entries, the
 557 quantity $\|\mathbf{A}\|_{\infty,1}$ is bounded.

558 Also, from (28), the expectation of $\|\tilde{\mathbf{Z}}\|_{2,2}^2$ is

$$\begin{aligned}\mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 &= \sum_{j=1}^K \sum_{i \in \mathcal{I}_j} \mathbb{E}(\langle \mathbf{a}[i], \mathbf{e} \rangle)^2 + 2\mathbb{E}z[i]\langle \mathbf{a}[i], \mathbf{e} \rangle \\ &\quad + \mathbb{E}z[i]^2 = \frac{KL}{m} \|\mathbf{A}\mathbf{e} + \mathbf{z}\|_2^2.\end{aligned}\tag{30}$$

559 Obtaining the the lower and upper bound of $\tilde{\Xi}$, we can then apply Hoeffding's inequality to get the
 560 tail bound of $\|\tilde{\mathbf{Z}}\|_{2,2}^2$. It can be written as follows: for any $\tau > 0$,

$$\begin{aligned}\mathbb{P}\{\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \tau \leq 0\} \\ \geq 1 - \exp \frac{-2\tau^2}{KL(\|\mathbf{A}\|_{\infty,1} \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^4}.\end{aligned}\tag{31}$$

561 Similarly, as in the proof of Theorem 4, here we consider the distance from the recovered solution
 562 $\hat{\mathbf{X}}$ to the exactly s -row-sparse solution \mathbf{X}_0 . Let $\tilde{\Delta}$ be $\tilde{\Delta} = \frac{1}{C_1(\delta)} \|\hat{\mathbf{X}} - \mathbf{X}_0\|_{2,2}$ and, according to
 563 Theorem 14, we have

$$\mathbb{P}\{\|\tilde{\Delta} - \|\tilde{\mathbf{Z}}\|_{2,2} \leq 0\} = 1.\tag{32}$$

564 Combing (31) and (32) allows us to conclude that

$$\begin{aligned}\mathbb{P}\{\tilde{\Delta}^2 - \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \tau^2 \leq 0\} \\ = \mathbb{P}\{\tilde{\Delta}^2 - \|\mathbf{Z}\|_{2,2}^2 + \|\tilde{\mathbf{Z}}\|_{2,2}^2 - \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \tau^2 \leq 0\} \\ \geq \mathbb{P}\{\tilde{\Delta}^2 - \|\mathbf{Z}\|_{2,2}^2 \leq 0, \|\tilde{\mathbf{Z}}\|_{2,2}^2 - \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \tau^2 \leq 0\} \\ = \mathbb{P}\{\tilde{\Delta}^2 - \|\mathbf{Z}\|_{2,2}^2 \leq 0\} \mathbb{P}\{\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 - \tau^2 \leq 0\} \\ \geq 1 - \exp \frac{-2\tau^4}{KL(\|\mathbf{A}\|_{\infty,1} \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^4}.\end{aligned}$$

565 We can then bound the expected square root of noise power:

$$\begin{aligned}\mathbb{P}\{\tilde{\Delta} \leq (\mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2)^{1/2} + \tau\} \quad (\text{by (21)}) \\ \geq \mathbb{P}\{\tilde{\Delta}^2 \leq \mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2 + \tau^2\} \\ \geq 1 - \exp \frac{-2\tau^4}{KL(\|\mathbf{A}\|_{\infty,1} \|\mathbf{e}\|_\infty + \|\mathbf{z}\|_\infty)^4}.\end{aligned}\tag{33}$$

566 The final JOBS estimates \mathbf{x}^J is $\mathbf{x}^J = \frac{1}{K} \sum_{j=1}^K \hat{\mathbf{x}}_j$ and as a direct result of (24), we have

$$\begin{aligned}
\|\mathbf{x}^J - \mathbf{x}^\star\|_2 &\leq \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}^\star\|_{2,2} \\
&\text{(by (27))} \\
&\leq \frac{1}{\sqrt{K}} \|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{2,2} + \|\mathbf{e}\|_2 = \frac{\mathcal{C}_1(\delta)\tilde{\Delta}}{\sqrt{K}} + \|\mathbf{e}\|_2.
\end{aligned} \tag{34}$$

567 Combing the results from (33) and (34) yields

$$\begin{aligned}
\mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 \leq \frac{\mathcal{C}_1(\delta)((\mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2)^{1/2} + \tau)}{\sqrt{K}} + \|\mathbf{e}\|_2\} \\
\geq \mathbb{P}\{\frac{\mathcal{C}_1(\delta)\tilde{\Delta}}{\sqrt{K}} + \|\mathbf{e}\|_2 \leq \frac{\mathcal{C}_1(\delta)((\mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2)^{1/2} + \tau)}{\sqrt{K}} + \|\mathbf{e}\|_2\} \\
= \mathbb{P}\{\tilde{\Delta} \leq (\mathbb{E}\|\tilde{\mathbf{Z}}\|_{2,2}^2)^{1/2} + \tau\} \\
\geq 1 - \exp \frac{-2k\tau^4}{(\|\mathbf{A}\|_{\infty,1}\|\mathbf{e}\|_{\infty} + \|\mathbf{z}\|_{\infty})^4}.
\end{aligned} \tag{35}$$

568 Finally, by substituting in the expected noise level derived in (30), we arrive at

$$\begin{aligned}
\mathbb{P}\{\|\mathbf{x}^J - \mathbf{x}^\star\|_2 \\
\leq \mathcal{C}_1(\delta)(\sqrt{\frac{L}{m}}\|\mathbf{A}\mathbf{e} + \mathbf{z}\|_2 + \frac{\tau}{\sqrt{K}}) + \|\mathbf{e}\|_2\} \\
\geq 1 - \exp \frac{-2\tau^4}{KL(\|\mathbf{A}\|_{\infty,1}\|\mathbf{e}\|_{\infty} + \|\mathbf{z}\|_{\infty})^4}.
\end{aligned} \tag{36}$$

569 Replacing τ with τ/\sqrt{K} , the quantity on the right hand side of the equation then becomes $1 -$
570 $\exp \frac{-2K\tau^4}{L(\|\mathbf{A}\|_{\infty,1}\|\mathbf{e}\|_{\infty} + \|\mathbf{z}\|_{\infty})^4}$ and we have proved the theorem.

571 9 Appendix D: Proofs of Bagging theorems

572 9.1 Proof of Bagging Performance bound for exactly s -sparse signals

573 Let $\mathbf{x}_1^B, \mathbf{x}_2^B, \dots, \mathbf{x}_K^B$ be the solutions of individually solved problems and the solution of the bagging
574 scheme \mathbf{x}^B is obtained from their average: $\mathbf{x}^B = \frac{1}{K} \sum_{j=1}^K \mathbf{x}_j^B$. We consider the distance to the true
575 solution \mathbf{x}^\star from each estimate separately. Here, the desired upper bound is the square root of the
576 expected power of each noise vector: $(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2} = \sqrt{\frac{L}{m}}\|\mathbf{z}\|_2$, where \mathcal{I} is a multi-set of size L
577 with each element randomly sampled from $\{1, 2, \dots, m\}$. For any $\tau > 0$, we have:

$$\begin{aligned}
&\mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - \mathcal{C}_1(\delta)((\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2} + \tau) \leq 0\} \\
&= \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - \mathcal{C}_1(\delta)((\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2} + \tau)^2 \leq 0\} \\
&\geq \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - \mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2 + \tau^2)^{1/2} \leq 0\} \\
&= \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2 + \tau^2) \leq 0\}.
\end{aligned}$$

578 Consider using the average of errors for each estimate $\frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2$, we can establish

$$\begin{aligned}
& \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - \mathcal{C}_1(\delta)((\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2} + \tau) \leq 0\} \\
&= \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \\
&+ \frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - \mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) \leq 0\} \\
&\geq \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \leq 0, \\
&\frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - \mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) \leq 0\}
\end{aligned}$$

579 (from the independence of two terms)

$$\begin{aligned}
&= \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \leq 0\} \\
&\times \mathbb{P}\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - K\mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) \leq 0\}.
\end{aligned}$$

580 By Jensen's inequality, the bagging error is smaller than the averaged error of each individual
581 estimator as in (23) and the first term holds with probability 1. Therefore, we have:

$$\begin{aligned}
& \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - \mathcal{C}_1(\delta)((\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2} + \tau) \leq 0\} \\
&\geq \mathbb{P}\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - K\mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) \leq 0\} \\
&= 1 - \mathbb{P}\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \geq K\mathcal{C}_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2)\}.
\end{aligned} \tag{37}$$

582 From this procedure, we can reduce the error bound for the bagging algorithm to bound the sum of
583 individual errors.

584 Let the random variable of error for each bagged estimator be $\mathbf{x}(\mathcal{I})$: $\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2$, where \mathcal{I} denotes
585 a bootstrap sample of size L and $\mathbf{x}(\mathcal{I})$ is the bagged solution from ℓ_1 minimization on the bootstrap
586 sample \mathcal{I} : $\mathbf{x}(\mathcal{I}) = \arg \min \|\mathbf{x}\|_1$ s.t. $\|\mathbf{y}_{[T]} - \mathbf{A}[\mathcal{I}]\|_2^2 \leq \epsilon(\mathcal{I})$. The power of all errors for each bagged
587 estimators $\|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2$ are realizations generated i.i.d. from the distribution of $\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2$. We
588 proceed with the proof using the following lemma that establishes the tail bound of the sum of i.i.d.
589 bounded random variables. It is a generalization of Hoeffding's inequality and the details of its proof
590 can be found in Appendix 10.4.

591 **Lemma 17 (Tail bound of the sum of i.i.d. bounded Random variables)** Let Y_1, Y_2, \dots, Y_n be
592 i.i.d. observations of bounded random variable Y : $a \leq Y \leq b$ and the expectation $\mathbb{E}Y$ exists.
593 Then, for any $\zeta > 0$,

$$\mathbb{P}\{\sum_{i=1}^n Y_i \geq n\zeta\} \leq \exp\{-\frac{2n(\zeta - \mathbb{E}Y)^2}{(b-a)^2}\}. \tag{38}$$

594 In this case, we consider the lower bound a and the upper bound b of the error $\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2$. Clearly
595 this term is non-negative, hence, we can set $a = 0$. The upper bound is obtained from the error bound
596 of ℓ_1 -minimization in Theorem 13. For all \mathcal{I} :

$$\mathbb{P}\{\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 - \mathcal{C}_1^2(\delta)\|\mathbf{z}_{[T]}\|_2^2 \leq 0\} = 1. \tag{39}$$

597 According to the norm equivalence inequality, we have

$$\|\mathbf{z}_{[T]}\|_2^2 \leq (\sqrt{L}\|\mathbf{z}_{[T]}\|_\infty)^2 \leq (\sqrt{L}\|\mathbf{z}\|_\infty)^2 = L\|\mathbf{z}\|_\infty^2. \tag{40}$$

598 From this, we can set $b = C_1^2(\delta)L\|\mathbf{z}\|_\infty^2$.

599 We can now apply the sum of i.i.d. bounded random variable in Theorem 17 to analyze our problem.
 600 By (37), the parameter ζ in (38) turns out to be: $\zeta = C_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2)$. Hence,

$$\begin{aligned} \mathbb{P}\left\{\sum_{j=1}^K \|\mathbf{x}_j - \mathbf{x}^\star\|_2^2 - K\zeta \geq 0\right\} &\leq \\ &\exp\left\{-\frac{2K(\zeta - \mathbb{E}\|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2)}{C_1^4(\delta)L^2\|\mathbf{z}\|_\infty^4}\right\}. \end{aligned} \quad (41)$$

601 To simplify the right hand side, let us consider the expected bagged error: $\mathbb{E}\|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2 =$
 602 $\frac{1}{|m^L|} \sum_{\mathcal{I}} \|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2$. Our bound in (39) implies that

$$\mathbb{P}\left\{\frac{1}{|m^L|} \sum_{\mathcal{I}} \|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2 \leq \frac{1}{|m^L|} \sum_{\mathcal{I}} C_1^2(\delta)\|\mathbf{z}_{\mathcal{I}}\|_2^2\right\} = 1,$$

603 which is equivalent to

$$\begin{aligned} \mathbb{E}\|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2 &\leq \frac{1}{|m^L|} \sum_{\mathcal{I}} C_1^2(\delta)\|\mathbf{z}_{\mathcal{I}}\|_2^2 \\ &= \mathbb{E} C_1^2(\delta)\|\mathbf{z}_{\mathcal{I}}\|_2^2 = C_1^2(\delta)\mathbb{E}\|\mathbf{z}_{\mathcal{I}}\|_2^2. \end{aligned} \quad (42)$$

604 From here, it is easy to see that

$$\begin{aligned} &\zeta - \mathbb{E}\|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2 \\ &= C_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) - \mathbb{E}\|\mathbf{x}_{(T)} - \mathbf{x}^\star\|_2^2 \\ &\geq C_1^2(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2 + \tau^2) - C_1^2(\delta)\mathbb{E}\|\mathbf{z}_{\mathcal{I}}\|_2^2 = C_1^2(\delta)\tau^2. \end{aligned} \quad (43)$$

605 The right hand side of (41) is upper bounded by $\exp\left\{-\frac{2K\tau^4}{L^2\|\mathbf{z}\|_\infty^4}\right\}$. Substituting this result into (37),
 606 we can obtain the result in our main bagging theorem.

607 9.2 Proof of Bagging performance bound of bagging for approximately sparse signals

608 In this section, we are working with the case when the true solution \mathbf{x}^\star is only approximately sparse.
 609 In other words, its sparsity level may exceed s and the s -sparse approximation error is no longer
 610 necessarily zero. Let ϵ_s denote the sparse approximation error $\epsilon_s = C_0(\delta)s^{-1/2}\|\mathbf{e}\|_1$. The square
 611 root of the expected power of each noise vector is $(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2} = \sqrt{\frac{L}{m}}\|\mathbf{z}\|_2$. We consider the
 612 following bound:

$$\begin{aligned} &\mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2 - (\epsilon_s + C_1(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2} + \tau) \leq 0\} \\ &= \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - (\epsilon_s + C_1(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2} + \tau)^2 \leq 0\} \\ &\geq \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \\ &\quad ((\epsilon_s + C_1(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2})^2 + C_1^2(\delta)\tau^2) \leq 0\}. \end{aligned}$$

613 Set $\zeta' = (\epsilon_s + C_1(\delta)(\mathbb{E}\|\mathbf{z}_{[T]}\|_2^2)^{1/2})^2 + C_1^2(\delta)\tau^2$ and consider using the averages of the errors
 614 $\frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2$ as an intermediate term. Repeating the same proving technique as in (37)
 615 yields

$$\begin{aligned} \mathbb{P}\{\|\mathbf{x}^B - \mathbf{x}^\star\|_2^2 - \zeta'\} &\geq \mathbb{P}\left\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - K\zeta' \leq 0\right\} \\ &= 1 - \mathbb{P}\left\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \geq K\zeta'\right\}. \end{aligned}$$

616 According to Lemma 17, we have:

$$\mathbb{P}\left\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 \geq K\zeta'\right\} \leq \exp\left\{-\frac{2K(\zeta' - \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2)^2}{(b' - a')^2}\right\}. \quad (44)$$

617 Here, $a' = 0$ and $b' = (\epsilon_s + \mathcal{C}_1(\delta)\sqrt{L}\|\mathbf{z}\|_\infty)^2$. The lower bound a' is set to zero since the error
618 for any bagged estimator $\|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2$ is non-negative. The upper bound b' can be obtained using
619 Theorem 13 and substituting in the upper bound of the noise power as derived in (40).

620 Next, consider the term $\zeta' - \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 = (\mathcal{C}_0(\delta)s^{-1/2}\|\mathbf{e}\|_1 + \mathcal{C}_1(\delta)\sqrt{\frac{L}{m}}\|\mathbf{z}\|_2)^2 + \mathcal{C}_1^2(\delta)\tau^2 -$
621 $\mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2$. We can upper bound the expected value of the error of bagged estimator with same
622 approach in (42). From Theorem 13, for all \mathcal{I} :

$$\mathbb{P}\{\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 \leq (\epsilon_s + \mathcal{C}_1(\delta)\|\mathbf{z}[\mathcal{I}]\|_2)^2\} = 1. \quad (45)$$

623 Since \mathcal{I} takes value of all m^L choices with equal probability, the following result is implied from
624 (45):

$$\mathbb{P}\{\mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 \leq \mathbb{E}(\epsilon_s + \mathcal{C}_1(\delta)\|\mathbf{z}[\mathcal{I}]\|_2)^2\} = 1. \quad (46)$$

625 Since $f(x) = x^2$ is a convex function, applying Jensen's inequality results in

$$(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2)^2 \leq \mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2.$$

626 Since the square root $x^{1/2}$ is a increasing function of x , taking square root preserves the sign of the
627 inequality:

$$\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2 \leq (\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2}. \quad (47)$$

628 Then, from (46), we have:

$$\begin{aligned} \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 &\leq \mathbb{E}(\epsilon_s + \mathcal{C}_1(\delta)\|\mathbf{z}[\mathcal{I}]\|_2)^2 \\ &= \epsilon_s^2 + \mathcal{C}_1^2(\delta)\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2 + 2\epsilon_s\mathcal{C}_1(\delta)\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2 \\ &\quad (\text{by (47)}) \\ &\leq \epsilon_s^2 + \mathcal{C}_1^2(\delta)\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2 + 2\epsilon_s\mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2} \\ &= (\epsilon_s + \mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2})^2. \end{aligned}$$

629 Finally, we can bound the term $\zeta' - \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2$:

$$\begin{aligned} &\zeta' - \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 \\ &= (\epsilon_s + \mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2})^2 + \mathcal{C}_1^2(\delta)\tau^2 - \mathbb{E}\|\mathbf{x}(\mathcal{I}) - \mathbf{x}^\star\|_2^2 \\ &\geq ((\epsilon_s + \mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2})^2 + \mathcal{C}_1^2(\delta)\tau^2 \\ &\quad - (\epsilon_s + \mathcal{C}_1(\delta)(\mathbb{E}\|\mathbf{z}[\mathcal{I}]\|_2^2)^{1/2})^2) = \mathcal{C}_1^2(\delta)\tau^2. \end{aligned}$$

630 One can observe that the upper bound of this difference is $\mathcal{C}_1^2(\delta)\tau^2$, which is the same as in the case
631 of the exact s -sparse signal in (43). The bound for (44) can be upper bounded by

$$\mathbb{P}\left\{\sum_{j=1}^K \|\mathbf{x}_j^B - \mathbf{x}^\star\|_2^2 - K\zeta' \geq 0\right\} \leq \exp\left\{-\frac{2K\mathcal{C}_1^4(\delta)\tau^4}{(b')^2}\right\},$$

632 where $b' = (\mathcal{C}_0(\delta)s^{-1/2}\|\mathbf{e}\|_1 + \mathcal{C}_1(\delta)\sqrt{L}\|\mathbf{z}\|_\infty)^2$.

633 10 Appendix E: Theory for NSP and RIP

634 10.1 Proof of the reverse direction for noiseless recovery

635 **Lemma 18** *If the MMV problem $\mathbf{P}_1(K)$, $K > 1$, in (56) has a unique solution, it will be of form*
636 $\mathbf{X}^\star = (\mathbf{x}^\star, \mathbf{x}^\star, \dots, \mathbf{x}^\star)$. *Then, there is a unique solution to \mathbf{P}_1 : \mathbf{x}^\star .*

Let us prove the other direction. If $\mathbf{P}_1(K)$ has a unique solution, the solution must be in the form of $\mathbf{X}^* = (\mathbf{x}^*, \mathbf{x}^*, \dots, \mathbf{x}^*)$, and it implies that \mathbf{P}_1 has a unique solution \mathbf{x}^* .

If $\mathbf{P}_1(K)$ has a unique solution, then it is equivalent to say that \mathbf{A} satisfied BNSP of order s . For all $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K) \neq \mathbf{O}, \mathbf{v}_j \in \text{Null}(\mathbf{A})$, we have $\forall \mathcal{S}, |\mathcal{S}| \leq s, \|\mathbf{V}[\mathcal{S}]\|_{1,2} < \|\mathbf{V}[\mathcal{S}^c]\|_{1,2}$. This implies that $\forall \mathbf{V} = (\mathbf{v}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0}), \mathbf{v} \in \text{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, BNSP is satisfied. Since in this case, except the first column, all others are zero and therefore do not contribute any to the group norm. Mathematically, for all $\mathcal{S}, \|\mathbf{V}[\mathcal{S}]\|_{1,2} = \|\mathbf{v}[\mathcal{S}]\|_1$. We, therefore, will have the BNSP of order s , implying the NSP for \mathbf{A} of order s .

10.2 JOBS matrix satisfies BNSP implies that each block matrix satisfies NSP

Using a similar analysis as in previous subsection 10.1, we conclude that a block diagonal matrix satisfies BNSP of order s implies that each submatrix satisfies NSP of order s . The block diagonal JOBS matrix $\mathbf{A}^J = \text{block_diag}(\mathbf{A}[\mathcal{I}_1], \mathbf{A}[\mathcal{I}_2], \dots, \mathbf{A}[\mathcal{I}_K])$ satisfies BNSP of order s . Then, for all $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K) \neq \mathbf{O}, \mathbf{v}_j \in \text{Null}(\mathbf{A}[\mathcal{I}_j]), j = 1, 2, \dots, K$, we have $\forall \mathcal{S}, |\mathcal{S}| \leq s, \|\mathbf{V}[\mathcal{S}]\|_{1,2} < \|\mathbf{V}[\mathcal{S}^c]\|_{1,2}$. This implies that $\forall \mathbf{V} = (\mathbf{0}, \dots, \mathbf{v}_j, \dots, \mathbf{0}), \mathbf{v}_j \in \text{Null}(\mathbf{A}[\mathcal{I}_j]) \setminus \{\mathbf{0}\}$, BNSP is satisfied, which essentially states that NSP is satisfied for $\mathbf{A}[\mathcal{I}_j]$.

10.3 Proof of Proposition 3

To prove this proposition, we give an alternative form of RIP and BRIP which are stated in the following two propositions. Alternative form of RIP as a function of matrix induced norm is given as follows.

Proposition 19 (Alternative form of RIP) Matrix \mathbf{A} has ℓ_2 -normalized columns, and $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathcal{S} \subset \{1, 2, \dots, n\}$ with size smaller or equal to s and $\mathbf{A}_{\mathcal{S}}$ takes columns of \mathbf{A} with indices in \mathcal{S} . The RIP constant of order s of \mathbf{A} , $\delta_s(\mathbf{A})$ is:

$$\delta_s(\mathbf{A}) = \max_{\mathcal{S} \subseteq \{1, 2, \dots, n\}, |\mathcal{S}| \leq s} \|\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} - \mathbf{I}\|_{2 \rightarrow 2}, \quad (48)$$

where \mathbf{I} is an identity matrix of size $s \times s$ and $\|\cdot\|_{2 \rightarrow 2}$ is the induced 2-norm defined as for any matrix \mathbf{A} , $\|\mathbf{A}\|_{2 \rightarrow 2} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$.

This proposition can be directly derived from the definition of RIP constant. Similarly, we can derive the alternative form of BRIP constant as a function of matrix induced norm.

Proposition 20 (Alternative form of BRIP) Let matrix \mathbf{A} have ℓ_2 -normalized columns and let $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n\}$ be the group sparsity pattern that defines the row sparsity pattern, with \mathcal{B}_i contains all indices corresponding to all elements of the i -th row. For $\mathcal{S} \subseteq \{1, 2, \dots, n\}$, denote $\mathcal{B}(\mathcal{S}) = \{\mathcal{B}_i, i \in \mathcal{S}\}$ as the subsets that takes several groups with group indices in \mathcal{S} . For $\mathbf{A} \in \mathbb{R}^{m \times n}$ with Block-RIP constant of order s , $\delta_{s|\mathcal{B}}(\mathbf{A})$ is

$$\delta_{s|\mathcal{B}} = \max_{\mathcal{S} \subseteq \{1, 2, \dots, n\}, |\mathcal{S}| \leq s} \|\mathbf{A}_{\mathcal{B}(\mathcal{S})}^T \mathbf{A}_{\mathcal{B}(\mathcal{S})} - \mathbf{I}\|_{2 \rightarrow 2}. \quad (49)$$

Without loss of generality, let us assume that all columns of \mathbf{A} in the original ℓ_1 minimization have unit ℓ_2 norms. Therefore, \mathbf{A} does not have any zero column. Before we calculate the RIP constant of the bootstrapped sensing matrices, we need to perform two operations: remove the duplicate rows from bootstrapped sensing matrices and then normalize the columns.

First, we remove the duplicated rows using the weighted scheme. In the noisy recovery problem, for a multi-set \mathcal{I} that may contain duplicate, the set \mathcal{U} denotes the set of all unique elements. In the constraint optimization, we can express the sum using occurrence times in \mathcal{I} for each element using c_i . $\|\mathbf{A}[\mathcal{I}]\mathbf{x} - \mathbf{y}[\mathcal{I}]\|_2^2 = \sum_{i \in \mathcal{I}} \|\mathbf{a}[i]\mathbf{x} - \mathbf{y}[i]\|_2^2 = \sum_{i \in \mathcal{U}} \|\sqrt{c_i}\mathbf{a}[i]\mathbf{x} - \sqrt{c_i}\mathbf{y}[i]\|_2^2$. Therefore, the original program is equivalent to reducing the duplicated rows in the bootstrap sample using $\sqrt{c_i}$ as weights. Because sampling with replacement is uniform, therefore the expected values of occurrence times for each sample are the same. To denote this operation, we have $\mathbf{R} \in \mathbb{R}^{u \times L}$, $\mathbf{R} = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \dots, \sqrt{c_u})\mathbf{I}[\mathcal{U}]$, each row of $\mathbf{I}[\mathcal{U}]$ corresponds to the unique vector of a row and this operation deletes the duplicated rows.

681 Second, we normalize the columns of these matrices using the following normalization procedure.
 682 For $M \in \mathbb{R}^{u \times n}$, since the original matrix A does not have any zero column, $Q(M) \in \mathbb{R}^{n \times n}$
 683 is a normalization matrix of M such that $MQ(M)$ has ℓ_2 -normalized columns. Clearly, the
 684 normalization matrix Q of M is obtained by:

$$Q(M) = \text{diag}(\|m_1\|_2^{-1}, \|m_2\|_2^{-1}, \dots, \|m_n\|_2^{-1}), \quad (50)$$

685 where m_j denotes j -th column of M .

686 Similary, we can construct Q_j s using (50) to normalize the columns. Let the original JOBS matrix
 687 be $A^J = \text{block_diag}(A[\mathcal{I}_1], A[\mathcal{I}_2], \dots, A[\mathcal{I}_K])$. We first normalize each block and then obtain the
 688 normalized bootstrapped sensing matrix as: $\widetilde{A}[\mathcal{I}_j] = R_j A[\mathcal{I}_j] Q_j$. The original JOBS matrix can be
 689 transferred into the normalized version $\widetilde{A}^J = \text{block_diag}(\widetilde{A}[\mathcal{I}_1], \widetilde{A}[\mathcal{I}_2], \dots, \widetilde{A}[\mathcal{I}_K])$.

690 Now, we consider the BRIP constant for A^J . In this derivation, column selection of a matrix is
 691 written as a right multiplication of the matrix $I_S(\cdot)$.

$$\begin{aligned} \delta_{s|B}(A^J) &= \delta_{s|B}(\widetilde{A}^J) \\ &= \max_{S \subseteq \{1, 2, \dots, n\}, |S| \leq s} \|(\widetilde{A}^J I_{B(S)})^T \widetilde{A}^J I_{B(S)} - I\|_{2 \rightarrow 2} \\ &= \max_{\substack{S \subseteq \{1, 2, \dots, n\}, \\ |S| \leq s}} \max_j \|(\widetilde{A}[\mathcal{I}_j] I_S)^T \widetilde{A}[\mathcal{I}_j] I_S - I\|_{2 \rightarrow 2} \\ &= \max_{\substack{S \subseteq \{1, 2, \dots, n\}, \\ |S| \leq s}} \|\text{block_diag}((\widetilde{A}[\mathcal{I}_1] I_S)^T \widetilde{A}[\mathcal{I}_1] I_S - I, \\ &\quad \dots, (\widetilde{A}[\mathcal{I}_K] I_S)^T \widetilde{A}[\mathcal{I}_K] I_S - I)\|_{2 \rightarrow 2}. \end{aligned}$$

692 The induced 2-norm of a matrix equals to the max singular value of $\|D\|_{2 \rightarrow 2} = \sigma_{\max}(D)$ and if D
 693 is a block diagonal matrix $D = \text{diag}(D_1, D_2, \dots, D_K)$, then $\sigma_{\max}(D) = \max_{j=1, 2, \dots, K} \sigma_{\max}(D_j)$.
 694 Applying this property leads to

$$\begin{aligned} \delta_{s|B}(A^J) &= \max_{\substack{S \subseteq \{1, 2, \dots, n\}, \\ |S| \leq s}} \max_j \|(\widetilde{A}[\mathcal{I}_j] I_S)^T \widetilde{A}[\mathcal{I}_j] I_S - I\|_{2 \rightarrow 2} \\ &= \max_j \max_{\substack{S \subseteq \{1, 2, \dots, n\}, \\ |S| \leq s}} \|(\widetilde{A}[\mathcal{I}_j] I_S)^T \widetilde{A}[\mathcal{I}_j] I_S - I\|_{2 \rightarrow 2} \\ &= \max_j \delta_s(A[\mathcal{I}_j]). \end{aligned}$$

695 10.4 Proof of Lemma 17

696 To prove of this lemma, We would need the Markov's inequality for non-negative random variables
 697 here. Let X be a non-negative random variable and suppose that $\mathbb{E}X$ exists. For any $t > 0$, we have:

$$\mathbb{P}\{X > t\} \leq \frac{\mathbb{E}X}{t}. \quad (51)$$

698 We also need the upper bound of the moment generating function (MGF) of the random variable Y .
 699 Suppose that $a \leq Y \leq b$, then for all $t \in \mathbb{R}$,

$$\mathbb{E} \exp\{tY\} \leq \exp\{t\mathbb{E}Y + \frac{t^2(b-a)^2}{8}\}. \quad (52)$$

700 Back to Lemma 17, for $t > 0$,

$$\begin{aligned} \mathbb{P}\left\{\sum_{i=1}^n Y_i \geq n\zeta\right\} &= \mathbb{P}\left\{\exp\left\{\sum_{i=1}^n Y_i\right\} \geq \exp\{n\zeta\}\right\} \\ &= \mathbb{P}\left\{\exp\left\{t \sum_{i=1}^n Y_i\right\} \geq \exp\{tn\zeta\}\right\} \end{aligned}$$

701 using the Markov inequality in (51)

$$\begin{aligned}
&\leq \exp\{-tn\zeta\} \mathbb{E}\{\exp\{t \sum_{i=1}^n Y_i\}\} \\
&= \exp\{-tn\zeta\} \mathbb{E}\{\Pi_{i=1}^n \exp\{tY_i\}\} \\
&= \exp\{-tn\zeta\} \Pi_{i=1}^n \mathbb{E}\{\exp\{tY_i\}\}
\end{aligned}$$

702 by upper bound for MGF in (52)

$$\begin{aligned}
&\leq \exp\{-tn\zeta\} (\exp\{t\mathbb{E}Y + \frac{t^2(b-a)^2}{8}\})^n \\
&= \exp\{-tn\zeta + tn\mathbb{E}Y + \frac{t^2(b-a)^2n}{8}\}.
\end{aligned}$$

703 The right hand side is a convex function with respect to t . Taking the derivative with respect to t and
704 set it zero, we obtain the optimal t , $t^* = \frac{4\zeta - 4\mathbb{E}Y}{(b-a)^2}$. The right hand side is minimized at value:

$$\exp\{-t^*n\zeta + t^*n\mathbb{E}Y + \frac{t^{*2}(b-a)^2n}{8}\} = \exp\{\frac{-2n(\zeta - \mathbb{E}Y)^2}{(b-a)^2}\}.$$

705 11 Appendix F: Pseudo-code of JOBS implementation via ADMM

706 We present the pseudo-code for solving JOBS optimization problem via ADMM updates. The key
707 difference to Bagging and the baseline ℓ_1 minimization here is that we employ the soft-thresholding
708 operation on each row in JOBS (described in line 6 of Algorithm 1), rather than the common
709 entry-wise thresholding operation on each individual sparse-code element in Bagging.

Algorithm 1 ADMM for solving JOBS

Require: Sensing matrix and measurements vector (\mathbf{A}, \mathbf{y}) , bootstrap ratio and number of estimates $(L/m, K)$, sparse balancing ratio λ , learning rate ρ , maximum number of iterations MaxIter.

Initialization: $\widehat{\mathbf{X}}_0, \mathbf{W}_0, \mathbf{U}_0 \leftarrow \mathbf{O}$ (zero matrix of size $n \times K$)

1: generate K bootstrap samples of length L :

$\{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_K\}$, and its corresponding $\{\mathbf{A}[\mathcal{I}_j], \mathbf{y}[\mathcal{I}_j]\}$

2: **for** $t = 1 : \text{MaxIter}$ **do**

3: $\widehat{\mathbf{X}}$ update: $\widehat{\mathbf{x}}_j \leftarrow$

$$(\mathbf{A}[\mathcal{I}_j]^* \mathbf{A}[\mathcal{I}_j] + \rho \mathbf{I})^{-1} (\mathbf{A}[\mathcal{I}_j]^* \mathbf{y}[\mathcal{I}_j] + \rho(\mathbf{w} - \mathbf{u}))$$

4: $\widehat{\mathbf{X}} \leftarrow \alpha \widehat{\mathbf{X}} + (1 - \alpha) \mathbf{W}$

5: \mathbf{W} update: applying shrinkage operations on each row. For $i = 1, 2, \dots, n$,

6: $\mathbf{w}[i] \leftarrow \text{Shrinkage}_{\lambda/\rho}(\widehat{\mathbf{x}}[i] - \mathbf{u}[i])$,

$$\text{Shrinkage}_{\kappa}(\mathbf{x}) = \max(1 - \kappa/\|\mathbf{x}\|_2, 0)\mathbf{x}$$

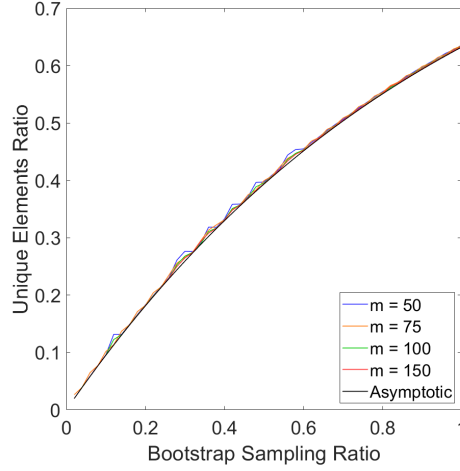
7: \mathbf{U} update: $\mathbf{U} = \mathbf{U} + \mathbf{X} - \mathbf{W}$

8: **end for**

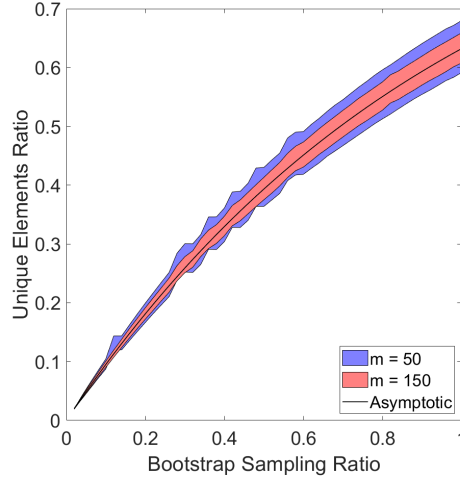
9: JOBS solution is the average columns of solution matrix $\widehat{\mathbf{X}}$: $\mathbf{x}^J = 1/K \sum \widehat{\mathbf{x}}_j$

710 12 Appendix G: Distribution of the unique number of elements for 711 bootstrapping

712 The bootstrap is essentially sampling with replacement, which is likely to create duplicate information.
713 The performance of sampling with replacement and sampling without replacement (sub-sampling)
714 can be linked by studying the quantity of the number of unique elements. In this section, we give the
715 analytic form of the number of unique samples when there are finite number of measurements m and
716 bootstrap sample L , as well as the form for asymptotic case as $m \rightarrow \infty$. The finite case is studied
717 in a well-known statistics problem – the Birthday Problem [23]. We also show empirically that the
718 finite m case is close in the asymptotic sense.



(a) The mean of unique element ratios under various sampling with replacement/ bootstrapping ratios with various number of measurements: $m = 50, 75, 100, 150$ and theoretical value when $m = \infty$.



(b) The area between of empirical mean plus and minus one empirical standard deviation. The blue area and the red area corresponds to the number of total measurements $m = 50$ and 150 respectively. The black line is the asymptotic mean and the asymptotic variance converges to zero

Figure 2: Unique element ratios with various bootstrapping ratios.

12.0.1 Unique number of Bootstrap Samples with Finite Sample m

We generate L samples from m samples uniformly at random with replacement ($L \leq m$). Let U denote the number of distinct samples among L samples. Clearly we have the number of distinct samples is between $[1, L]$ and the probability mass function is given by [23], same as the famous Birthday problem in statistics:

$$\mathbb{P}(U = u) = \binom{m}{u} \sum_{j=0}^u (-1)^j \binom{u}{j} \left(\frac{u-j}{m}\right)^L, \quad (53)$$

$$u = 1, 2, \dots, L.$$

724 In our problem, we are interested in finding the lower bound of U with certainty $1 - \alpha$

$$\mathbb{P}(U \geq d) = \sum_{U=d}^L \binom{m}{u} \sum_{j=0}^u (-1)^j \binom{u}{j} \left(\frac{u-j}{m}\right)^L \geq 1 - \alpha. \quad (54)$$

725 Therefore for

$$1 \geq \alpha \geq \sum_{u=0}^{d-1} \binom{m}{u} \sum_{j=0}^u (-1)^j \binom{u}{j} \left(\frac{u-j}{m}\right)^L, \quad (55)$$

726 equation (54) is satisfied.

727 12.0.2 Asymptotic unique ratios of bootstrap samples

728 The theoretically unique percentage for asymptotic case when the number of total number of mea-
729 surements goes to infinity $m \rightarrow \infty$ has been studied in the literature [24, 25]. In the limit case,
730 the limiting distribution of the number of unique elements U is normal. The asymptotic mean for
731 the unique number of elements over total number of measurements m is $\mathbb{E}\frac{U}{m} = 1 - \exp\{-r\}$,
732 where r is the bootstrap sampling rate. The asymptotic variance of the unique ratio is then
733 $\text{Var}\frac{U}{m} = \frac{1}{m}(\exp\{-r\} - (1+r)\exp\{-2r\})$, which converges to zero when m is large.

734 12.0.3 Finite number of measurements m cases are empirically close to the asymptotic case

735 We generate 10000 trials of random sampling with replacement and then calculate the empirical
736 unique percentage by counting the ratio of the number of unique elements over the total number of
737 measurements m . The theoretical mean is consistently lower than the mean for a finite m . From the
738 plot, the average unique elements in finite m cases $m = 50, 75, 100, 150$ are not so different from
739 the theoretical value of the infinite sample size.

740 The empirical mean and the asymptotic value are plotted in Figure 2a, indicating that the numeric
741 unique percentage is not that far from the asymptotic value even when the number of estimates is
742 finite and small. Figure 2b illustrates the region between the mean plus and minus one standard of
743 deviation. As the asymptotic case, the theoretical standard deviation converges to zero. We plotted
744 the cases $m = 150$ and $m = 50$ compared to the asymptotic case. For both, the variance is tight and
745 gets smaller when m becomes larger. For the same m , the variance of the unique number of elements
746 become larger when the bootstrap ratio L/m is large.

747 12.1 The sub-sampling Variation: Sub-JOBS

748 Bootstrapping (random sampling with replacement) creates duplicates within a bootstrap sample.
749 Although it simplifies the analysis, in practice, duplicate information does not add value. One natural
750 extension of the proposed framework is to use sub-sampling: sampling without replacement. The
751 sub-sampling variation of Bagging is known as Subagging estimator in the literature [26, 27]. We
752 adopt a similar name for the sub-sampling variation of the proposed method: Sub-JOBS. The only
753 difference to the original scheme is that for each bootstrap sample \mathcal{I}_j , L distinct samples are generated
754 by random sampling without replacement from m measurements.

755 In this paper, all the theoretical results are for the bootstrapping version for simplicity of presentation.
756 The numerical results and discussion for both the original bootstrapping scheme as well as the sub-
757 sampling variation will be shown in Section ?? . The connection between bootstrap and sub-sampling
758 is also explained in details in Appendix 12.

759 13 Appendix H: JOBS, a two step relaxation of ℓ_1 minimization

760 JOBS recovers the true sparse solution because it is essentially a relaxation of the original ℓ_1
761 minimization problem in a multiple vectors fashion. Therefore, it is not so surprising that JOBS
762 relaxation can recover the true solution: exactly in the noiseless case and within some neighbourhood
763 of the ground truth in noisy case.

764 We demonstrate that JOBS is a two-step relaxation procedure of ℓ_1 minimization. For a ℓ_1 mini-
 765 mization with a unique solution \mathbf{x}^* , the multiple measurement vectors (MMV) equivalence is: for
 766 $j = 1, 2, \dots, K$

$$\mathbf{P}_1(K) : \min \|\mathbf{X}\|_{1,1} \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}_j, \quad (56)$$

767 where $\|\mathbf{X}\|_{1,1} = \sum_i \|\mathbf{x}[i]^T\|_1$ as mentioned in Table 4. We show that this MMV form (56)
 768 is equivalent to the original ℓ_1 minimization problem. If the original problem \mathbf{P}_1 has a unique
 769 solution \mathbf{x}^* , then the solution to the MMV problem $\mathbf{P}_1(K)$ in (56) yields a row sparse solution
 770 $\mathbf{X}^* = (\mathbf{x}^*, \mathbf{x}^*, \dots, \mathbf{x}^*)$. This result can be derived via contradiction. The reverse direction is also
 771 true: if the MMV problem $\mathbf{P}_1(K)$ has a unique solution, it implies that the \mathbf{P}_1 must also have a
 772 unique solution. Details are stated in Lemma 18 in Appendix 10.1.

773 Since the $\ell_{1,1}$ norm of \mathbf{X} essentially takes ℓ_1 norm of its vectorized version, it only enforces the
 774 sparsity for all elements in \mathbf{X} without any structure such as the support consistency across its columns.
 775 To obtain the JOBS form, We first relax the $\ell_{1,1}$ norm in (56) to the $\ell_{1,2}$ norm. For all $j = 1, 2, \dots, K$

$$\mathbf{P}_{12}(K) : \min \|\mathbf{X}\|_{1,2} \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}_j. \quad (57)$$

776 From here, to obtain Noiseless JOBS version, we further drop all constraints that are not in \mathcal{I}_j from
 777 (57) for estimator $\mathbf{x}_j, j = 1, 2, \dots, K$. Then we obtain the noiseless version of JOBS:

$$\mathbf{J}_{12} : \min \|\mathbf{X}\|_{1,2} \text{ s.t. } \mathbf{y}_{[\mathcal{I}_j]} = \mathbf{A}_{[\mathcal{I}_j]}\mathbf{x}_j, \quad (58)$$

778 This two-step relaxation process is illustrated in Figure 3.

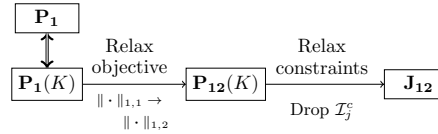


Figure 3: JOBS framework is a two-step relaxation of ℓ_1 minimization

779 The noisy version can be obtained similarly. We formulate the MMV version of the original ℓ_1
 780 problem; relax the regularizer from $\ell_{1,1}$ norm to $\ell_{1,2}$ norm, and then further relax the objective
 781 function by dropping the constraints that are not on the selected subset \mathcal{I}_j for the j -th estimate \mathbf{x}_j to
 782 obtain the proposed form \mathbf{J}_{12}^λ .

783 Because JOBS procedure is a two-step relaxation of the ℓ_1 minimization, it gives some insight of why
 784 JOBS algorithm can correctly recover sparse solution, which is important for analyzing the algorithm.
 785 In Section , we will establish the correctness of JOBS algorithm rigorously.