# Chapter 8. Sparse Recovery with Random Matrices

I-Liang Chern

May 12, 2016

# What's kind of matrices satisfying RIP

- Random matrices with
  - iid Gaussian entries
  - iid Bernoulli entries $(+/-1)$
  - iid subgaussian entries
  - random Fourier ensemble
  - random ensemble in bounded orthogonal systems
- In each case, $m = O(s \ln N)$, they satisfy RIP with very high probability $(1 - e^{-Cm})$..

This is a note from S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer 2013.

# Subgaussian random matrices

## Definition

Let $\mathbf{A}$ be $m \times N$ real valued random matrix.

- Bernoulli random matrix: if each entries of $\mathbf{A}$ are independent Rademacher random variables (i.e. takes $\pm 1$ with probability $1/2$ on each);
- Gaussian random matrix: if each entries of $\mathbf{A}$ are independent standard Gaussian random variables;
- Subgaussian random matrix: if each entries of $\mathbf{A}$ are independent subgaussian random variables with mean $0$ and variance $1$, and satisfying

$$P(|a_{jk}| \geq t) \leq \beta e^{-\kappa t^2}, \quad \text{for all } j \in [m], k \in [N].$$

**Remark.** Bernoulli random matrices and Gaussian random matrices are subgaussian random matrices.

# RIP for Subgaussian random matrices

## Theorem (Main theorem)

*Let* $\mathbf{A}$ *be an* $m \times N$ *subgaussian random matrix. Then there exists a constant* $C > 0$ *(depending on the subgaussian parameters* $\beta$ *and* $\kappa$*) such that for any* $0 < \delta < 1$ *the restricted isometry property of* $\mathbf{A}/\sqrt{m}$ *satisfies* $\delta_s < \delta$ *with probability at least* $1 - 2\exp(-\delta^2 m/(2C))$ *provided*

$$m \geq 2C\delta^{-2} s \ln(eN/s).$$

**Remark.** The term $1/\sqrt{m}$ is due to that we want to have the normalization of column vectors. Since the variance of each entries is $1$, the variance of the column vector $E[|a_j|^2] = m$. Thus, $E[|a_j/\sqrt{m}|^2] = 1$. This normalization yields

$$E[\| \frac{1}{\sqrt{m}} \mathbf{A}\mathbf{x} \|^2] = \|\mathbf{x}\|^2.$$

# Theorem

*Let* $\mathbf{A}$ *be an* $m \times N$ *subgaussian random matrix. Then there exists a constant* $C > 0$ *(depending on the subgaussian parameters* $\beta$ *and* $\kappa$*) and universal constants* $D_1$ *and* $D_2$ *such that if*

$$m \geq 2C\delta^{-2} s \ln(eN/s)$$

*then the following statement holds with probability at least* $1 - 2\exp(-\delta^2 m/(2C))$*, uniformly for every* $\mathbf{x} \in \Sigma_s$*: given* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *with* $\|\mathbf{e}\|_2 \leq \sqrt{m}\eta$*, for some* $\eta > 0$*, a solution* $\mathbf{x}^\#$ *of*

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_1 \quad \textit{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \sqrt{m}\eta$$

*satisfies the estimates*

$$\|\mathbf{x} - \mathbf{x}^\#\|_2 \leq D_1 \frac{\sigma_s(\mathbf{x})_1}{\sqrt{s}} + D_2\eta$$
$$\|\mathbf{x} - \mathbf{x}^\#\|_1 \leq D_1 \sigma_s(\mathbf{x})_1 + D_2\sqrt{s}\eta.$$

Proof. The above optimization problem is equivalent to

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{z} - \frac{1}{\sqrt{m}}\mathbf{y}\|_2 \leq \eta.$$

We apply the main theorem to $\frac{1}{\sqrt{m}}\mathbf{A}$ and get this theorem.

# Outline of the Proof

1. A concentration lemma: Let $\mathbf{A}$ be a $m \times N$ subgaussian matrix with subgaussian parameter $c > 0$. Then for any $\mathbf{x} \in \mathbb{R}^N$ and any $t \in (0, 1)$,

$$P(|m^{-1}\|\mathbf{Ax}\|^2 - \|\mathbf{x}\|^2| \geq t\|\mathbf{x}\|^2) \leq 2\exp(-\tilde{c}mt^2),$$

   where $\tilde{c}$ depends only on $c$.

2. Part 1. given $S \subset [N]$ with $|S| = s$ and given $\delta, \epsilon \in (0, 1)$, if $m > C\delta^{-2}(7s + 2\ln(2\epsilon^{-1}))$, where $C = 2/(3\tilde{c})$, then with probability at least $1 - \epsilon$,

$$\|\mathbf{A}_S^*\mathbf{A}_S - Id\|_{2\to 2} < \delta.$$

3. Part II. for $\delta, \epsilon \in (0, 1)$ and when

$$m \geq C\delta^{-2}[s(9 + 2\ln(N/s)) + 2\ln(2\epsilon^{-1})],$$

   where $C = 2/(3\tilde{c})$, then $\delta_s(\mathbf{A}) < \delta$ with probability at least $1 - \epsilon$.

Proof concentration lemma.

1. We denote $\mathbf{A} = (Y_1^t, ..., Y_m^t)^t$, where $Y_\ell^t$, $\ell = 1, ..., m$ are the row vectors of $\mathbf{A}$. For any $\mathbf{x} \in \mathbb{R}^N$, let $Z_\ell := |\langle Y_\ell, \mathbf{x}\rangle|^2 - \|\mathbf{x}\|^2$. Then

$$m^{-1}\|\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2 = \frac{1}{m}\sum_{\ell=1}^{m}(|\langle Y_\ell, \mathbf{x}\rangle|^2 - \|\mathbf{x}\|^2) = \frac{1}{m}\sum_{\ell=1}^{m} Z_\ell$$

2. For each $\ell$, from the independence of $a_{\ell,j}$ and $a_{\ell,k}$, we have

$$E|\langle Y_\ell, \mathbf{x}\rangle|^2 = \sum_{j,k=1}^{N} x_j x_k E[a_{\ell,j}a_{\ell,k}] = \sum_{j=1}^{N} x_j^2 = \|\mathbf{x}\|^2.$$

   Thus, $E[Z_\ell] = 0$ for all $\ell \in [m]$.

3. Since each $a_{\ell,k}$ is subgaussian, we have $\langle Y_\ell, \mathbf{x}\rangle = \sum_{j=1}^{N} x_j a_{\ell,j}$ is also subgaussian. Thus, $Z_\ell = |\langle Y_\ell, \mathbf{x}\rangle|^2$ is subexponential. That is, there exist $\beta, \kappa > 0$ such that

$$P(|Z_\ell| \geq r) \leq \beta \exp(-\kappa r).$$

4. From Bernstein inequality for subexponential random variables,

$$P\left(|\frac{1}{m}\sum_{\ell=1}^{m} Z_\ell| \geq t\right) = P\left(|\sum_{\ell=1}^{m} Z_\ell| \geq tm\right) \leq 2\exp\left(-\frac{\kappa^2 m^2 t^2/2}{2\beta m + \kappa m t}\right)$$

$$\leq 2\exp\left(-\frac{\kappa^2}{4\beta + 2\kappa} m t^2\right) = 2\exp(-\tilde{c} m t^2).$$

Proof of Part I.

1. Let us consider the ball

$$B_S := \{\mathbf{x} \,|\, \text{supp}\, \mathbf{x} \subset S, \|\mathbf{x}\| \le 1\}.$$

We will cover $B_S$ by $B_S \subset \bigcup_{u \in U} B_\rho(u)$, where the radius $\rho \in (0, 1/2)$ is to be chosen later. The center set $U$ are chosen as the follows. On the coordinate axis $x_i$ for $i \in S$, we choose $u = k\rho \mathbf{e}_i$, $|k| \le 1/\rho$. Thus, there are at most $1 + 2/\rho$ such centers on $x_i$ coordinate axis. The center set $U$ is chosen to be the Cartesian product of such centers from each coordinate axis $x_i$ with $i \in S$. Thus,

$$|U| \le \left(1 + \frac{2}{\rho}\right)^s.$$

2. The concentration inequality gives

$$\begin{aligned}
&P(|\|\mathbf{A}u\|^2 - \|u\|^2| \ge t\|u\|^2 \text{ for some } u \in U) \\
&\le \sum_{u \in U} P(|\|\mathbf{A}u\|^2 - \|u\|^2| \ge t\|u\|^2) \le 2|U| \exp(-\tilde{c}mt^2) \\
&\le 2\left(1 + \frac{2}{\rho}\right)^s \exp(-\tilde{c}mt^2).
\end{aligned}$$

This is equivalent to

$$P(|\|\mathbf{A}u\|^2 - \|u\|^2| < t\|u\|^2 \text{ for all } u \in U) > 1 - 2\left(1 + \frac{2}{\rho}\right)^s \exp(-\tilde{c}mt^2).$$

3 Now, for any $\mathbf{x} \in B_S$, there exists $u \in U$ such that $\|\mathbf{x} - u\| \leq \rho < 1/2$. Let us write $A_S^* A_S - Id$ by $B$.

$$|\langle B\mathbf{x}, \mathbf{x}\rangle| = |\langle Bu, u\rangle + \langle B(\mathbf{x}+u), \mathbf{x}-u\rangle| \leq |\langle Bu, u\rangle| + |\langle B(\mathbf{x}+u), \mathbf{x}-u\rangle|$$
$$< t + \|B\|_{2\to 2}\|\mathbf{x}+u\|\|\mathbf{x}-u\| \leq t + 2\rho\|B\|_{2\to 2}.$$

This gives

$$\|B\|_{2\to 2} < t + 2\rho\|B\|_{2\to 2}, \quad \text{i.e.} \quad \|B\|_{2\to 2} \leq \frac{t}{1-2\rho}.$$

4 Now, we choose $t = (1-2\rho)\delta$. Then

$$P(\|A_S^* A_S - Id\|_{2\to 2} \geq \delta) \leq 2\left(1 + \frac{2}{\rho}\right)^s \exp(-\tilde{c}m(1-2\rho)^2\delta^2)$$

In order to have the right hand side to be smaller than $\epsilon$, it requires

$$m \geq \frac{1}{\tilde{c}(1-2\rho)^2}\delta^{-2}(\ln(1+2/\rho)s + \ln(2\epsilon^{-1})).$$

Now we choose $\rho = 2/(e^{7/2}-1) \approx 0.0623$ so that $1/(1-2\rho)^2 \leq 4/3$ and $\ln(1+2/\rho)/(1-2\rho)^2 \leq 14/3$. We get

$$P(\|A_S^* A_S - Id\|_{2\to 2} \geq \delta) \leq \epsilon$$

when

$$m \geq \frac{2}{3\tilde{c}}\delta^{-2}(7s + 2\ln(2\epsilon^{-1})).$$

Proof of Part II.

1. The restricted isometry constant

$$\delta_s = \sup_{S\subset[N],|S|=s} \|A_S^* A_S - Id\|_{2\to 2}.$$

There are $\binom{N}{s}$ such $S$. Thus

$$\begin{aligned} P(\delta_s \geq \delta) &\leq \sum_{S\subset[N],|S|=s} P(\|A_S^* A_S - Id\|_{2\to 2} \geq \delta) \\ &\leq 2\binom{N}{s}\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}\delta^2(1-2\rho)^2 m\right) \\ &\leq 2\left(\frac{eN}{s}\right)^s\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}\delta^2(1-2\rho)^2 m\right) \end{aligned}$$

Here, we have used

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leq \frac{n^k}{k!} = \frac{k^k n^k}{k! k^k} \leq e^k \frac{n^k}{k^k}.$$

2 Taking $\rho = 2/(e^{7/2} - 1)$ as before yields that $\delta_s < \delta$ with probability at least $1 - \epsilon$ provided

$$2\left(\frac{eN}{s}\right)^s \left(1 + \frac{2}{\rho}\right)^s \exp\left(-\tilde{c}\delta^2(1-2\rho)^2 m\right) \leq \epsilon$$

This equivalent to

$$m \geq \frac{1}{\tilde{c}\delta^2}\left(\frac{4}{3}s\ln(eN/s) + \frac{14}{3}s + \frac{4}{3}\ln(2\epsilon^{-1})\right) \geq C\delta^{-2}(s\ln(eN/s) + \ln(2\epsilon^{-1})).$$

where $C$ only depends on $\tilde{c}$.

3 Taking $\epsilon = 2\exp(-\delta^2 m/(2C))$ yields that the condition $m \geq 2C\delta^{-2}s\ln(eN/s)$ guarantees that $\delta_s \leq \delta$ with probability at least $1 - 2\exp(-m\delta^2/(2C))$.

# Relation to the Johnson-Linderstrauss lemma

This concentration inequality is closely related to the classical John-Lindenstrauss embedding lemma. We can also say that the J-L lemma leads to RIP.

## Lemma (Johnson-Linderstrauss)

*Let* $\mathbf{x}_1, ..., \mathbf{x}_M$ *be* $M$ *points in* $\mathbb{R}^N$*. Let* $\eta \in (0,1)$ *be a constant. Then there exists a universal constant* $C > 0$ *and an* $m \times N$ *matrix* $\mathbf{B}$ *such that if* $m > C\eta^{-2} \ln M$*, then*

$$(1-\eta)\|\mathbf{x}_j - \mathbf{x}_\ell\|_2^2 \le \|\mathbf{B}(\mathbf{x}_j - \mathbf{x}_\ell)\|_2^2 \le (1+\eta)\|\mathbf{x}_j - \mathbf{x}_\ell\|_2^2$$

*for any* $j, \ell \in [M]$*.*

1. Considering the set $E = \{\mathbf{x}_j - \mathbf{x}_\ell \mid, j, \ell \in [M],\ j < \ell\}$. We have $|E| = M(M-1)/2$.
2. Let $\mathbf{A}$ be an $m \times N$ subgaussian random matrix. By concentration inequality, there exists a $\tilde{c} > 0$ such that for each $\mathbf{x} \in E$, we have

$$P(|\|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2| > \eta\|\mathbf{x}\|^2) \leq 2\exp(-\tilde{c}m\eta^2).$$

   Thus,

$$P(|\|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2| \leq \eta\|\mathbf{x}\|^2 \text{ for all } \mathbf{x} \in E) \geq 1 - M^2\exp(-\tilde{c}m\eta^2).$$

3. For $\epsilon \in (0,1)$, if we choose $m \geq \tilde{c}^{-1}\eta^{-2}\ln(M/\epsilon)$, then $M^2\exp(-\tilde{c}m\eta^2) < \epsilon$. Since $\epsilon < 1$,

$$P(|\|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2| \leq \eta\|\mathbf{x}\|^2 \text{ for all } \mathbf{x} \in E) \geq 1 - \epsilon > 0.$$

   We can recursively generate subgaussian matrix $\mathbf{A}$ until

$$|\|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2| \leq \eta\|\mathbf{x}\|^2 \text{ for all } \mathbf{x} \in E$$

   is valid. Such random matrix exists because the probability this happens is greater than $0$.

# Nonuniform recovery via random subgaussian matrices

## Theorem

*Let* $\mathbf{x} \in \mathbb{C}^N$ *be an* $s$*-sparse vector. Let* $\mathbf{A} \in \mathbb{R}^{m \times N}$ *be a subgaussian random matrix with subgaussian parameter* $c$*. If, for some* $\epsilon \in (0, 1)$,

$$m \geq \frac{4c}{1-\delta} s \ln(2N/\epsilon), \text{ with } \delta = \sqrt{\frac{C}{4c}\left(\frac{7}{\ln(2N/\epsilon)} + \frac{2}{s}\right)}$$

*(assuming* $N$ *and* $s$ *are large enough so that* $\delta < 1$*), then with probability at least* $1 - \epsilon$ *the vector* $\mathbf{x}$ *is the unique minimizer of*

$$\|\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}.$$

*The constant* $C = 2/(3\tilde{c})$ *depends only on the subgaussian parameter c.*

# Remark: Difference between uniform and nonuniform recovery

- Uniform recovery guarantee provides a lower probability estimate of the form

$$P(\forall\ s\text{-sparse } \mathbf{x}, \text{ recovery of } \mathbf{x} \text{ is successful using } \mathbf{A}) \geq 1 - \epsilon.$$

- Nonuniform recovery gives a statement of the form

$$\forall s\text{-sparse vector } \mathbf{x},\ \ P(\text{ recovery of } \mathbf{x} \text{ is successful using } \mathbf{A}) \geq 1 - \epsilon.$$

# What's kind of matrices satisfying RIP

- Random matrices with
  - iid Gaussian entries
  - iid Bernoulli entries $(+/-1)$
  - iid subgaussian entries
  - random Fourier ensemble
  - random ensemble in bounded orthogonal systems
- In each case, $m = O(s \ln N)$, they satisfy RIP with very high probability $(1 - e^{-Cm})$..

This is a note from S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer 2013.

# Random Sampling from Bounded Orthonormal Systems

- In applications, we would like to represent data in terms of good basis such as Fourier, wavelets. Suppose our function is sparse in terms of some orthonormal basis. The question is how to sample them in order to recover the function exactly.
- Let $\mathcal{D} \subset \mathbb{R}^d$ be endowed with a probability measure $\nu$. A set $\{\phi_1, ..., \phi_N\}$ defined on $\mathcal{D}$ is called a bounded orthonormal system (BOS) if
  - $\int_D \phi_i(t)\overline{\phi_j(t)}\, d\nu(t) = \delta_{ij}$
  - There exists a $K > 0$ such that $\|\phi_j\|_\infty \leq K$ for all $j \in [N]$.

# Random Sampling from BOS-2

- We consider function $f$ defined on $\mathcal{D}$ of the form

$$f(t) = \sum_{i=1}^{N} x_i \phi_i(t).$$

- Let $t_1, ..., t_m \in \mathcal{D}$ be sampling points, chosen randomly according to the probability $\nu$ on $\mathcal{D}$.
- We are given the sampled values

$$y_\ell = f(t_\ell) = \sum_{k=1}^{N} \phi_k(t_\ell) x_k, \quad \ell \in [m].$$

The matrix $\mathbf{A} := (\phi_k(t_\ell)),\ \ell \in [m],\ k \in [N]$ is called the randomly sampling matrix associated with the BOS with bound $K$.

# Examples of BOS

- Trigonometric polynomials. Let $\mathcal{D} = [0, 1]$, $\nu$ the Lebesgue measure on $[0, 1]$,

  $$\phi_k(t) = e^{2\pi i k t}.$$

  We find the constant $K = 1$. We choose $\Gamma \subset \mathbb{Z}$ to be a set of size $N$. The set $\{\phi_k | k \in \Gamma\}$ is a BOS. The sampling points $t_1, ..., t_m$ are chosen independently and uniformly at random from $[0, 1]$.

# Examples of BOS

- Discrete orthonormal systems. Let $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_N] \in \mathbb{C}^N \times \mathbb{C}^N$ be a unitary matrix. The set $\mathcal{D} := [N]$ and the measure $\nu$ is the counting measure. The normalized function

$$\phi_k(t) : \sqrt{N}\mathbf{u}_k(t), \quad t \in [N].$$

The inner product of $\phi_k$ and $\phi_\ell$ is

$$\int \phi_k(t)\overline{\phi_\ell(t)}d\nu(t) = \frac{1}{N}\sum_{t=1}^{N} \sqrt{N}\mathbf{u}_k(t)\overline{\sqrt{N}\mathbf{u}_\ell(t)} = \langle \mathbf{u}_k, \mathbf{u}_\ell \rangle = \delta_{k\ell}.$$

We should require that there exists $K > 0$ such that

$$\sqrt{N} \max_{t,k\in[N]} |\mathbf{u}_k(t)| \leq K.$$

We choose $t_1, ..., t_m$ independently, uniformly at random from $[N]$. The randomly sampled matrix $\mathbf{A}$ is defined to be $\mathbf{A} = \mathbf{R}_T\sqrt{N}\mathbf{U}$, where $T = \{t_1, ..., t_m\}$ and $\mathbf{R}_T : \mathbb{C}^N \to \mathbb{C}^m$: $(\mathbf{R}_T\mathbf{z})_\ell = z_{t_\ell}, \quad \ell \in [m]$. As a concrete example is the discrete Fourier transform

$$\mathbf{U}_{k\ell} = \frac{1}{\sqrt{N}}e^{2\pi i(k-1)(\ell-1)/N}, \quad k, \ell \in [N].$$

# Examples of BOS

- Hadamard matrix. Hadamard matrix $H_n \in \mathbb{R}^{2^n \times 2^n}$ can be viewed as a Fourier transform on $\mathbb{Z}_2^n = \{0,1\}^n$. For $j, \ell \in [2^n]$, write them in binary expansion

$$j = \sum_{k=1}^{n} j_k 2^{k-1} + 1, \quad \ell = \sum_{k=1}^{n} \ell_k 2^{k-1} + 1,$$

where $j_k, \ell_k \in \{0,1\}$. The Hadamard matrix $H_n$ is defined as

$$H_{j,\ell} = \frac{1}{2^{n/2}} (-1)^{\sum_{k=1}^{n} j_k \ell_k}.$$

The Hadamard matrix has a recursive expression

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \quad H_0 = (1),$$

which leads to a fast matrix-vector multiplication. The Hadamard matrix is self-adjoint and orthogonal

$$H_n = H_n^* = H_n^{-1}.$$

Its column vectors form a BOS with $K = 1$.

# Nonuniform recovery via random sampling BOS

A sequence $\epsilon = (\epsilon_1, ..., \epsilon_N)$ is called a Rademacher sequence if $\epsilon_i$ are independent Rademacher variable (i.e. it takes values $\pm 1$ with probability $1/2$.)

A complex random variable which is uniformly distributed on the torus

$\{z \in \mathbb{C} | |z| = 1\}$ is called a Steinhaus variable. A sequence $\epsilon = (\epsilon_1, ..., \epsilon_N)$ of

independent Steinhaus variables is called a Steinhaus sequence.

# Nonuniform recovery via random sampling BOS

## Theorem

*Let* $\mathbf{x} \in \mathbb{C}^N$ *be a vector supported on a set* $S$ *of size* $s$ *such that sign*$(x_S)$ *forms a Rademacher or Steinhaus sequence. Let* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *be a random sampling matrix associated with a BOS with bound* $K \geq 1$*. If*

$$m \geq CK^2 s \ln^2(6N/\epsilon),$$

*then with probability at least* $1 - \epsilon$*, the vector* $\mathbf{x}$ *is the unique minimizer of*

$$\min \|\mathbf{z}\|_1 \quad \textit{subject to} \quad \mathbf{Az} = \mathbf{Ax}.$$

*Here* $\epsilon \in (0, 1)$ *and the constant* $C \leq 35$*.*

# Uniform recovery via random sampling BOS

## Theorem

*Let* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *be a random sampling matrix associated to a BOS with constant* $K \geq 1$*. If, for* $\delta \in (0,1)$*,*

$$m \geq CK^2\delta^{-2}s\ln^4(N),$$

*then with probability at least* $1 - N^{-\ln^3(N)}$ *the restricted isometry constant* $\delta_s$ *of* $\frac{1}{\sqrt{m}}\mathbf{A}$ *satisfies* $\delta_s \leq \delta$*. The constant* $C > 0$ *is universal.*

# Uniform robust recovery via random sampling BOS

## Corollary

*Let* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *be a random sampling matrix associated to a BOS with constant* $K \geq 1$*. Suppose that*

$$m \geq CK^2 s \ln^4(N)$$

*for a universal constant* $C > 0$*. Then with probability at least* $1 - N^{-\ln^3(N)}$

(a) *every* $s$*-sparse vector* $\mathbf{x}$ *is exactly recovered from* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *by basis pursuit;*

(b) *every* $s$*-sparse vector* $\mathbf{x}$ *is approximately recovered from the inaccurate samples* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, $\|\mathbf{e}\|_2 \leq \sqrt{m}\eta$*, as a solution* $\mathbf{x}^\#$ *of*

$$\min \|\mathbf{z}\|_1 \quad \textit{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \sqrt{m}\eta$$

*in the sense that*

$$\|\mathbf{x}^\# - \mathbf{x}\|_p \leq \frac{C_1}{s^{1-1/p}} \sigma_s(\mathbf{x})_1 + C_2 s^{1/p-1/2}\eta, \quad 1 \leq p \leq 2$$

*where the constant* $C_1, C_2$ *are universal.*