Compressive Sensing

Structured Random Matrices

Holger Rauhut
Department of Mathematics
RWTH Aachen University

Winter School on Statistics
Borgafjöll March 4 – 8, 2013
Why structure?

- Applications impose structure due to physical constraints, limited freedom to inject randomness.
- Fast matrix vector multiplies (FFT) in recovery algorithms, unstructured random matrices impracticable for large scale applications.
- Storage problems for unstructured matrices.
Why structure?

- Applications impose structure due to physical constraints, limited freedom to inject randomness.
- Fast matrix vector multiplies (FFT) in recovery algorithms, unstructured random matrices impracticable for large scale applications.
- Storage problems for unstructured matrices.
Why structure?

- Applications impose structure due to physical constraints, limited freedom to inject randomness.
- Fast matrix vector multiplies (FFT) in recovery algorithms, unstructured random matrices impracticable for large scale applications.
- Storage problems for unstructured matrices.

Here

- Random sampling in bounded orthonormal systems (in particular, random partial Fourier matrices)
- Partial random circulant matrices.
Bounded orthonormal systems (BOS)

\( \mathcal{D} \subset \mathbb{R}^d \) endowed with probability measure \( \nu \).

\( \psi_1, \ldots, \psi_N : \mathcal{D} \to \mathbb{C} \) function system on \( \mathcal{D} \).
Bounded orthonormal systems (BOS)

\( \mathcal{D} \subset \mathbb{R}^d \) endowed with probability measure \( \nu \).

\( \psi_1, \ldots, \psi_N : \mathcal{D} \to \mathbb{C} \) function system on \( \mathcal{D} \).

Orthonormality

\[
\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k} = \begin{cases} 
0 & \text{if } j \neq k, \\
1 & \text{if } j = k.
\end{cases}
\]
Bounded orthonormal systems (BOS)

$\mathcal{D} \subset \mathbb{R}^d$ endowed with probability measure $\nu$.

$\psi_1, \ldots, \psi_N : \mathcal{D} \to \mathbb{C}$ function system on $\mathcal{D}$.

Orthonormality

$$\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k} = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$$

Uniform bound in $L^\infty$:

$$\|\psi_j\|_\infty = \sup_{t \in \mathcal{D}} |\psi_j(t)| \leq K \quad \text{for all } j \in [N].$$
Bounded orthonormal systems (BOS)

\( \mathcal{D} \subset \mathbb{R}^d \) endowed with probability measure \( \nu \).

\( \psi_1, \ldots, \psi_N : \mathcal{D} \to \mathbb{C} \) function system on \( \mathcal{D} \).

Orthonormality

\[
\int_{\mathcal{D}} \psi_j(t)\overline{\psi_k(t)}d\nu(t) = \delta_{j,k} = \begin{cases} 
0 & \text{if } j \neq k, \\
1 & \text{if } j = k.
\end{cases}
\]

Uniform bound in \( L^\infty \):

\[
\|\psi_j\|_\infty = \sup_{t \in \mathcal{D}} |\psi_j(t)| \leq K \quad \text{for all } j \in [N].
\]

Observe that \( K \geq 1 \):

\[
1 = \int_{\mathcal{D}} |\psi_j(t)|^2 d\nu(t) \leq \sup_{t \in \mathcal{D}} |\psi_j(t)|^2 \int_{\mathcal{D}} 1d\nu(t) \leq K^2.
\]
Consider functions

\[ f(t) = \sum_{k=1}^{N} x_k \psi_k(t), \quad t \in \mathcal{D}. \]
Consider functions

\[ f(t) = \sum_{k=1}^{N} x_k \psi_k(t), \quad t \in \mathcal{D}. \]

\( f \) is called \( s \)-sparse if \( x \) is \( s \)-sparse.
Consider functions

\[ f(t) = \sum_{k=1}^{N} x_k \psi_k(t), \quad t \in \mathcal{D}. \]

\( f \) is called \( s \)-sparse if \( x \) is \( s \)-sparse.

Sampling points \( t_1, \ldots, t_m \in \mathcal{D} \). Sample values:

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

\[ f(t) = \sum_{k=1}^{N} x_k \psi_k(t), \quad t \in \mathcal{D}. \]

\( f \) is called \( s \)-sparse if \( x \) is \( s \)-sparse.

Sampling points \( t_1, \ldots, t_m \in \mathcal{D} \). Sample values:

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

Sampling matrix \( A \in \mathbb{C}^{m \times N} \) with entries

\[ A_{\ell,k} = \psi_k(t_\ell), \quad \ell \in [m], k \in [N]. \]
Consider functions

\[ f(t) = \sum_{k=1}^{N} x_k \psi_k(t), \quad t \in \mathcal{D}. \]

\( f \) is called \( s \)-sparse if \( x \) is \( s \)-sparse.

Sampling points \( t_1, \ldots, t_m \in \mathcal{D} \). Sample values:

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

Sampling matrix \( A \in \mathbb{C}^{m \times N} \) with entries

\[ A_{\ell,k} = \psi_k(t_\ell), \quad \ell \in [m], k \in [N]. \]

Then

\[ y = Ax. \]
Problem: Reconstruct $s$-sparse $f$ — equivalently $x$ — from its sample values $y = Ax$. 
Problem: Reconstruct \( s \)-sparse \( f \) — equivalently \( x \) — from its sample values \( y = Ax \).

Behavior of \( A \) as measurement matrix?
Choose sampling points $t_1, \ldots, t_\ell$ independently at random according to the measure $\nu$, that is,

$$\mathbb{P}(t_\ell \in B) = \nu(B), \quad \text{for all measurable } B \subset \mathcal{D}.$$
Random Sampling

Choose sampling points \( t_1, \ldots, t_\ell \) independently at random according to the measure \( \nu \), that is,

\[
\mathbb{P}(t_\ell \in B) = \nu(B), \quad \text{for all measurable } B \subset \mathcal{D}.
\]

The sampling matrix \( A \) is then a structured random matrix.
Examples of Bounded Orthonormal Systems

**Trigonometric System.** $\mathcal{D} = [0, 1]$ with Lebesgue measure.

$$\psi_k(t) = e^{2\pi i k t}, \quad t \in [0, 1].$$

The trigonometric system is orthonormal with $K = 1$. 
Examples of Bounded Orthonormal Systems

**Trigonometric System.** $\mathcal{D} = [0, 1]$ with Lebesgue measure.

$$\psi_k(t) = e^{2\pi ikt}, \quad t \in [0, 1].$$

The trigonometric system is orthonormal with $K = 1$. Take subset $\Gamma \subset \mathbb{Z}$ of cardinality $N$; trigonometric polynomials

$$f(t) = \sum_{k \in \Gamma} x_k \psi_k(t) = \sum_{k \in \Gamma} x_k e^{2\pi ikt}.$$
**Examples of Bounded Orthonormal Systems**

**Trigonometric System.** $\mathcal{D} = [0, 1]$ with Lebesgue measure.

$$\psi_k(t) = e^{2\pi i kt}, \quad t \in [0, 1].$$

The trigonometric system is orthonormal with $K = 1$. Take subset $\Gamma \subset \mathbb{Z}$ of cardinality $N$; trigonometric polynomials

$$f(t) = \sum_{k \in \Gamma} x_k \psi_k(t) = \sum_{k \in \Gamma} x_k e^{2\pi i kt}.$$  

Samples $t_1, \ldots, t_m$ are chosen independently and uniformly at random from $[0, 1]$. 
Trigonometric System. $\mathcal{D} = [0, 1]$ with Lebesgue measure.

\[ \psi_k(t) = e^{2\pi i kt}, \quad t \in [0, 1]. \]

The trigonometric system is orthonormal with $K = 1$. Take subset $\Gamma \subset \mathbb{Z}$ of cardinality $N$; trigonometric polynomials

\[ f(t) = \sum_{k \in \Gamma} x_k \psi_k(t) = \sum_{k \in \Gamma} x_k e^{2\pi i kt}. \]

Samples $t_1, \ldots, t_m$ are chosen independently and uniformly at random from $[0, 1]$. Nonequispaced random Fourier matrix

\[ A_{\ell, k} = e^{2\pi i k t_\ell}, \quad \ell \in [m], k \in \Gamma. \]
Examples of Bounded Orthonormal Systems

**Trigonometric System.** \( \mathcal{D} = [0, 1] \) with Lebesgue measure.

\[
\psi_k(t) = e^{2\pi ikt}, \quad t \in [0, 1].
\]

The trigonometric system is orthonormal with \( K = 1 \). Take subset \( \Gamma \subset \mathbb{Z} \) of cardinality \( N \); trigonometric polynomials

\[
f(t) = \sum_{k \in \Gamma} x_k \psi_k(t) = \sum_{k \in \Gamma} x_k e^{2\pi ikt}.
\]

Samples \( t_1, \ldots, t_m \) are chosen independently and uniformly at random from \([0, 1]\). Nonequispaced random Fourier matrix

\[
A_{\ell, k} = e^{2\pi ikt_\ell}, \quad \ell \in [m], k \in \Gamma.
\]

Fast matrix vector multiply using the nonequispaced fast Fourier transform (NFFT).
Fourier-Coefficients
Fourier-Coefficients

Time-Domain Signal with 16 Samples
Fourier-Coefficients

Time-Domain Signal with 16 Samples

Traditional Reconstruction
Fourier-Coefficients

Time-Domain Signal with 16 Samples

Traditional Reconstruction

Reconstruction via $\ell_1$-minimization
**Real trigonometric polynomials.** $\mathcal{D} = [0, 1]$ with Lebesgue measure.

\[
\psi_{2k}(t) = \sqrt{2} \cos(2\pi kt), \quad k \in \mathbb{N}_0, \quad \psi_0(t) = 1, \\
\psi_{2k+1}(t) = \sqrt{2} \sin(2\pi kt), \quad k \in \mathbb{N}.
\] (1)

They form an orthonormal system on $[0, 1]$ with respect to the Lebesgue measure; $K = \sqrt{2}$. 
**Real trigonometric polynomials.** $\mathcal{D} = [0, 1]$ with Lebesgue measure.

\[
\psi_{2k}(t) = \sqrt{2} \cos(2\pi kt), \quad k \in \mathbb{N}_0, \quad \psi_0(t) = 1,
\]
\[
\psi_{2k+1}(t) = \sqrt{2} \sin(2\pi kt), \quad k \in \mathbb{N}.
\] (1)

They form an orthonormal system on $[0, 1]$ with respect to the Lebesgue measure; $K = \sqrt{2}$.
The samples $t_1, \ldots, t_m$ are chosen according to the uniform distribution on $[0, 1]$. 
$U = (U_{t,k}) \in \mathbb{C}^{N \times N}$ unitary matrix. Set $D = [N]$ and

$$\psi_k(t) = \sqrt{N} U_{t,k}, \quad k, t \in [N].$$
Discrete orthonormal systems

\[ U = (U_{t,k}) \in \mathbb{C}^{N \times N} \text{ unitary matrix. Set } \mathcal{D} = [N] \text{ and } \]

\[ \psi_k(t) = \sqrt{N} U_{t,k}, \quad k, t \in [N]. \]

Orthonormality with respect to discrete uniform probability measure \( \nu(B) = |B|/N, \ B \subset [N], \)

\[ \frac{1}{N} \sum_{t=1}^{N} \psi_k(t) \overline{\psi_j(t)} = (U^* U)_{k,j} = \delta_{k,j}. \]
Discrete orthonormal systems

\[ U = (U_{t,k}) \in \mathbb{C}^{N \times N} \text{ unitary matrix. Set } D = [N] \text{ and } \]

\[ \psi_k(t) = \sqrt{N} U_{t,k}, \quad k, t \in [N]. \]

Orthonormality with respect to discrete uniform probability measure \( \nu(B) = |B|/N, \ B \subset [N], \)

\[ \frac{1}{N} \sum_{t=1}^{N} \psi_k(t) \bar{\psi}_j(t) = (U^* U)_{k,j} = \delta_{k,j}. \]

Boundedness condition

\[ \sup_{j \in [N]} \sup_{t \in [N]} |\psi_k(t)| = \sqrt{N} \sup_{t, k \in [N]} |U_{t,k}| \leq K. \]
Discrete orthonormal systems

\[ U = (U_{t,k}) \in \mathbb{C}^{N \times N} \] unitary matrix. Set \( \mathcal{D} = [N] \) and

\[ \psi_k(t) = \sqrt{N} U_{t,k}, \quad k, t \in [N]. \]

Orthonormality with respect to discrete uniform probability measure \( \nu(B) = |B|/N, \ B \subset [N], \)

\[ \frac{1}{N} \sum_{t=1}^{N} \psi_k(t) \overline{\psi_j(t)} = (U^* U)_{k,j} = \delta_{k,j}. \]

Boundedness condition

\[ \sup_{j \in [N]} \sup_{t \in [N]} |\psi_k(t)| = \sqrt{N} \sup_{t,k \in [N]} |U_{t,k}| \leq K. \]

Samples are taken independent and uniformly on \([N]\). Sampling matrix \( A \) is a submatrix of \( U \) consisting of randomly chosen rows.
Random Partial Fourier Matrix

Take \( U = F \), Fourier matrix with entries

\[
F_{\ell,k} = \frac{1}{\sqrt{N}} e^{2\pi i \ell k / N}, \quad \ell, k \in [N].
\]

\( F \) is unitary and \( K = 1 \).
Take $U = F$, Fourier matrix with entries

$$F_{\ell,k} = \frac{1}{\sqrt{N}} e^{2\pi i \ell k / N}, \quad \ell, k \in [N].$$

$F$ is unitary and $K = 1$.

$\hat{x} = Fx$ is called Fourier transform of $x$. 

Random Partial Fourier Matrix
Take $U = F$, Fourier matrix with entries

$$F_{\ell,k} = \frac{1}{\sqrt{N}} \exp \frac{2\pi i \ell k}{N}, \quad \ell, k \in [N].$$

$F$ is unitary and $K = 1$. 

$\hat{x} = Fx$ is called Fourier transform of $x$.

Random sampling in this context corresponds to randomly subsampling the entries of $\hat{x}$. The resulting random sampling matrix is called random partial Fourier matrix.
Take $U = F$, Fourier matrix with entries

$$F_{\ell,k} = \frac{1}{\sqrt{N}} e^{2\pi i \ell k / N}, \quad \ell, k \in [N].$$

$F$ is unitary and $K = 1$.
$\hat{x} = Fx$ is called Fourier transform of $x$.

Random sampling in this context corresponds to randomly subsampling the entries of $\hat{x}$. The resulting random sampling matrix is called random partial Fourier matrix.

$F$, and hence $A$, has a fast matrix vector multiply, the famous fast Fourier transform (FFT).
Comparison of a traditional MRI reconstruction (left) and a compressive sensing reconstruction (right). Acquisition accelerated by a factor of 7.2 by random subsampling of the frequency domain.

Image courtesy of Michael Lustig, Stanford University, and Shreyas Vasanawala, Lucile Packard Children’s Hospital, Stanford University.
Incoherent Bases

\[ V = (v_\ell), \ W = (w_\ell) \in \mathbb{C}^{N \times N}: \text{unitary matrices.} \]
Incoherent Bases

\[ V = (v_\ell), \quad W = (w_\ell) \in \mathbb{C}^{N \times N}: \text{unitary matrices.} \]

Let \( z \in \mathbb{C}^N \) be \( s \)-sparse with respect to \( V \), that is, \( z = V x \) for an \( s \)-sparse vector \( x \).
Incoherent Bases

\[ V = (v_{\ell}), \quad W = (w_{\ell}) \in \mathbb{C}^{N \times N}: \text{unitary matrices.} \]

Let \( z \in \mathbb{C}^N \) be \( s \)-sparse with respect to \( V \), that is, \( z = Vx \) for an \( s \)-sparse vector \( x \).

Sample \( z \) with respect to \( W \), i.e.

\[ y_k = \langle z, w_{t_k} \rangle = \langle Vx, w_{t_k} \rangle = \langle x, V^* w_{t_k} \rangle = (W^* Vx)_{t_k}, \quad k \in [m]. \]
Incoherent Bases

\( V = (v_\ell), \; W = (w_\ell) \in \mathbb{C}^{N \times N} \): unitary matrices.
Let \( z \in \mathbb{C}^N \) be \( s \)-sparse with respect to \( V \), that is, \( z = Vx \) for an \( s \)-sparse vector \( x \).
Sample \( z \) with respect to \( W \), i.e.

\[
y_k = \langle z, w_{t_k} \rangle = \langle Vx, w_{t_k} \rangle = \langle x, V^* w_{t_k} \rangle = (W^* Vx)_{t_k}, \quad k \in [m].
\]

Defining \( U = W^* V \) we are back to the setting of discrete orthonormal systems.
$V = (v_\ell), W = (w_\ell) \in \mathbb{C}^{N \times N}$: unitary matrices.

Let $z \in \mathbb{C}^N$ be $s$-sparse with respect to $V$, that is, $z = Vx$ for an $s$-sparse vector $x$.

Sample $z$ with respect to $W$, i.e.

$$y_k = \langle z, w_{t_k} \rangle = \langle Vx, w_{t_k} \rangle = \langle x, V^* w_{t_k} \rangle = (W^* V x)_{t_k}, \quad k \in [m].$$

Defining $U = W^* V$ we are back to the setting of discrete orthonormal systems.

Boundedness condition leads to mutual incoherence

$$\sqrt{N} \max_{\ell, k \in [N]} |\langle v_\ell, w_k \rangle| \leq K.$$
\( V = (v_\ell), W = (w_\ell) \in \mathbb{C}^{N \times N} \): unitary matrices.

Let \( z \in \mathbb{C}^N \) be \( s \)-sparse with respect to \( V \), that is, \( z = Vx \) for an \( s \)-sparse vector \( x \).

Sample \( z \) with respect to \( W \), i.e.

\[
y_k = \langle z, w_{tk} \rangle = \langle Vx, w_{tk} \rangle = \langle x, V^*w_{tk} \rangle = (W^*Vx)_{tk}, \quad k \in [m].
\]

Defining \( U = W^*V \) we are back to the setting of discrete orthonormal systems.

Boundedness condition leads to mutual incoherence

\[
\sqrt{N} \max_{\ell,k \in [N]} |\langle v_\ell, w_k \rangle| \leq K.
\]

Special case: Haar-Wavelets and Noiselets
• Legendre polynomials, Chebyshev polynomials, Jacobi polynomials
  • Rescaling trick: sampling with respect to Chebyshev measure
    \[ d\nu(t) = \pi^{-1} (1 - t^2)^{-1/2} dt \]
• Spherical harmonics
Nonuniform versus uniform recovery results

We distinguish between two types of recovery theorems for random matrices:

• Uniform Recovery: With high probability every $s$-sparse vector can be recovered using $A$:
  $$\Pr(\forall s\text{-sparse } \mathbf{x}, \text{recovery of } \mathbf{x} \text{ is successful using } A) \geq 1 - \varepsilon.$$

• Nonuniform Recovery: A fixed $s$-sparse $\mathbf{x}$ can be recovered with high probability using a random draw of $A$:
  $$\forall s\text{-sparse } \mathbf{x}: \Pr(\text{recovery of } \mathbf{x} \text{ is successful using } A) \geq 1 - \varepsilon.$$

Uniform Recovery implies Nonuniform Recovery, but the converse does not hold. Non-uniform results are usually easier to show. Uniform results often use the RIP while non-uniform results are based on the recovery theorems for individual vectors.
Nonuniform versus uniform recovery results

We distinguish between two types of recovery theorems for random matrices:

- **Uniform Recovery**: With high probability every \(s\)-sparse vector can be recovered using \(A\):
  \[
  \mathbb{P}(\forall s\text{-sparse } x, \text{ recovery of } x \text{ is successful using } A) \geq 1 - \varepsilon.
  \]

- **Nonuniform Recovery**: A fixed \(s\)-sparse \(x\) can be recovered with high probability using a random draw of \(A\):
  \[
  \forall s\text{-sparse } x : \mathbb{P}(\text{recovery of } x \text{ is successful using } A) \geq 1 - \varepsilon.
  \]

Uniform Recovery implies Nonuniform Recovery, but the converse does not hold.
Nonuniform versus uniform recovery results

We distinguish between two types of recovery theorems for random matrices:

- **Uniform Recovery:** With high probability every $s$-sparse vector can be recovered using $A$:

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

- **Nonuniform Recovery:** A fixed $s$-sparse $\mathbf{x}$ can be recovered with high probability using a random draw of $A$:

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

Uniform Recovery implies Nonuniform Recovery, but the converse does not hold.

Non-uniform results are usually easier to show. Uniform results often use the RIP while non-uniform results are based on the recovery theorems for individual vectors.
Theorem

Let $\mathbf{x} \in \mathbb{C}^N$ be $s$-sparse. Choose $A \in \mathbb{C}^{m \times N}$ to be the random sampling matrix associated to a BOS with constant $K \geq 1$. Assume that

$$m \geq CK^2 s \ln(N) \ln(\varepsilon^{-1}).$$

Then with probability at least $1 - \varepsilon$ $\ell_1$-minimization recovers $\mathbf{x}$ from $\mathbf{y} = A\mathbf{x}$.

In the special case of the (discrete and continuous) Fourier system a slightly better condition is available:

$$m \geq Cs \ln(N/\varepsilon).$$
Restricted Isometry Property for Bounded Orthonormal Systems

Theorem (Candès, Tao ’06 – Rudelson, Vershynin ’06 – R ’08, ’09)

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

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

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

Refined condition

$$m \geq CK^2 \delta^{-2} s \max\{\ln^3(s) \ln(N), \ln^2(s) \ln(\ln(N)) \ln(N), \ln(\varepsilon^{-1})\}$$

for $\delta_s \leq \delta$ with probability at least $1 - \varepsilon$
**Theorem**

Let \( A \in \mathbb{C}^{m \times N} \) be the random sampling matrix associated to a BOS with constant \( K \geq 1 \). Let \( S \subset [N] \) be of cardinality \( |S| = s \). Then, for \( \delta \in (0, 1) \), the normalized matrix \( \tilde{A} = \frac{1}{\sqrt{m}} A \) satisfies

\[
\|\tilde{A}^*\tilde{A}S - I\|_{2 \rightarrow 2} \leq \delta \text{ with probability at least } 1 - 2s \exp \left( -\frac{3m\delta^2}{8K^2s} \right).
\]

Consequence: \( \tilde{A}_S \) is well-conditioned with probability at least \( 1 - \epsilon \) provided

\[
m \geq \frac{8}{3} K^2 \delta^{-2} s \ln(2s/\epsilon).
\]
Towards the nonuniform recovery result – Conditioning of submatrices

**Theorem**

Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$. Let $S \subset [N]$ be of cardinality $|S| = s$. Then, for $\delta \in (0, 1)$, the normalized matrix $\tilde{A} = \frac{1}{\sqrt{m}}A$ satisfies

$$\|\tilde{A}_S^*\tilde{A}_S - I\|_{2 \to 2} \leq \delta \text{ with probability at least } 1 - 2s \exp\left(-\frac{3m\delta^2}{8K^2s}\right).$$

Consequence: $\tilde{A}_S$ is well-conditioned with probability at least $1 - \epsilon$ provided

$$m \geq (8/3)K^2\delta^{-2}s \ln(2s/\epsilon).$$

Caveat: The union bound over all subset $S$ of cardinality $s$ does **not** give the stated RIP result!
Towards the nonuniform recovery result – Conditioning of submatrices

**Theorem**

Let \( A \in \mathbb{C}^{m \times N} \) be the random sampling matrix associated to a BOS with constant \( K \geq 1 \). Let \( S \subset [N] \) be of cardinality \( |S| = s \). Then, for \( \delta \in (0, 1) \), the normalized matrix \( \tilde{A} = \frac{1}{\sqrt{m}} A \) satisfies

\[
\|\tilde{A}_S^* \tilde{A}_S - I\|_2 \leq \delta \text{ with probability at least } 1 - 2s \exp \left( -\frac{3m\delta^2}{8K^2s} \right).
\]

Consequence: \( \tilde{A}_S \) is well-conditioned with probability at least \( 1 - \epsilon \) provided

\[
m \geq (8/3)K^2\delta^{-2}s \ln(2s/\epsilon).
\]

Caveat: The union bound over all subset \( S \) of cardinality \( s \) does not give the stated RIP result!
Theorem (Tropp '10)

Let \( X_1, \ldots, X_m \in \mathbb{C}^{d \times d} \) be independent mean-zero self-adjoint random matrices. Assume that

\[
\|X_\ell\|_{2 \rightarrow 2} \leq K \quad \text{almost surely,} \quad \ell \in [m],
\]

and set

\[
\sigma^2 := \left\| \sum_{\ell=1}^m \mathbb{E}(X^2_\ell) \right\|_{2 \rightarrow 2}.
\]

Then, for \( t > 0 \),

\[
P \left( \left\| \sum_{\ell=1}^m X_\ell \right\|_{2 \rightarrow 2} \geq t \right) \leq 2d \exp \left( -\frac{t^2/2}{\sigma^2 + Kt/3} \right).
\]

Circulant matrix: For $b = (b_0, b_1, \ldots, b_{N-1}) \in \mathbb{C}^N$ let

$\Phi = \Phi(b) \in \mathbb{C}^{N \times N}$ be the matrix with entries $\Phi_{i,j} = b_{j-i \mod N}$,

$$
\Phi(b) = \begin{pmatrix}
    b_0 & b_1 & \cdots & \cdots & b_{N-1} \\
    b_{N-1} & b_0 & b_1 & \cdots & b_{N-2} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    b_1 & b_2 & \cdots & b_{N-1} & b_0
\end{pmatrix}.
$$
Circulant matrix: For $b = (b_0, b_1, \ldots, b_{N-1}) \in \mathbb{C}^N$ let $\Phi = \Phi(b) \in \mathbb{C}^{N \times N}$ be the matrix with entries $\Phi_{i,j} = b_{j-i \mod N}$,

$$
\Phi(b) = \begin{pmatrix}
    b_0 & b_1 & \cdots & \cdots & b_{N-1} \\
    b_{N-1} & b_0 & b_1 & \cdots & b_{N-2} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    b_1 & b_2 & \cdots & b_{N-1} & b_0
\end{pmatrix}.
$$
Let $\Theta \subset [N]$ arbitrary of cardinality $m$.

$R_\Theta$: operator that restricts a vector $x \in \mathbb{C}^N$ to its entries in $\Theta$.

Restrict $\Phi(b)$ to the rows indexed by $\Theta$:

Partial circulant matrix: $\Phi^\Theta(b) = R_\Theta \Phi(b) \in \mathbb{C}^{m \times N}$
Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality $m$.

$R_{\Theta}$: operator that restricts a vector $x \in \mathbb{C}^N$ to its entries in $\Theta$. 

Restrict $\Phi(b)$ to the rows indexed by $\Theta$:

Partial circulant matrix: $\Phi^{\Theta}(b) = R_{\Theta}\Phi(b) \in \mathbb{C}^{m \times N}$
Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality $m$.

$R_{\Theta}$: operator that restricts a vector $x \in \mathbb{C}^N$ to its entries in $\Theta$.

Restrict $\Phi(b)$ to the rows indexed by $\Theta$:

Partial circulant matrix: $\Phi^\Theta(b) = R_{\Theta} \Phi(b) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:

$$y = R_{\Theta} \Phi(b) x = R_{\Theta} (b \ast x), \quad (b \ast x)_\ell = \sum_{j=1}^{N} b_{\ell-j \mod N} x_j$$
Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality $m$.

$R_\Theta$: operator that restricts a vector $x \in \mathbb{C}^N$ to its entries in $\Theta$.

Restrict $\Phi(b)$ to the rows indexed by $\Theta$:

Partial circulant matrix: $\Phi^\Theta(b) = R_\Theta \Phi(b) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:

$y = R_\Theta \Phi(b)x = R_\Theta(b \ast x)$, \hspace{1cm} (b \ast x)_\ell = \sum_{j=1}^{N} b_{\ell-j \mod N} x_j$

Matrix vector multiplication via the FFT!
Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality $m$.

$R_\Theta$: operator that restricts a vector $x \in \mathbb{C}^N$ to its entries in $\Theta$.

Restrict $\Phi(b)$ to the rows indexed by $\Theta$:

**Partial circulant matrix**: $\Phi_\Theta(b) = R_\Theta \Phi(b) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:

$$y = R_\Theta \Phi(b)x = R_\Theta (b \ast x), \quad (b \ast x)_\ell = \sum_{j=1}^{N} b_{\ell - j \mod N} x_j$$

Matrix vector multiplication via the FFT!

We choose the vector $b \in \mathbb{C}^N$ at random, in particular, as Rademacher sequence $b = \epsilon$, that is, $\epsilon_\ell = \pm 1$. 

Holger Rauhut, RWTH Aachen  
Compressive Sensing 22
Motivation I: Radar

Received signal is superposition of delayed versions of sent signal.

Task: Determine delays (corresponding to distances) from subsampled receive signal!

Advantage: uses slow analog-to-digital converters, avoids matched filter, can resolve small targets
Super-resolution based on compressive sensing

- Use coded mask instead of pinhole (or lense)
- Observed coded aperture image is subsampled $2D$-convolution of image $x$ with point-spread function $b$

Marcia, Willett 2009 – Romberg 2009
Numerical experiments

Sparse recovery via $\ell_1$-minimization with partial random circulant matrix $A \in \mathbb{R}^{m \times N}$, $N = 500$, $m = 100$. 

![Empirical Recovery Rate vs Sparsity graph]

Empirical Recovery Rate

Sparsity
Theorem (Rauhut 2009)

Let $\Theta \subset [N]$ be an arbitrary (deterministic) set of cardinality $m$. Let $x \in \mathbb{C}^N$ be $s$-sparse such that the signs of its non-zero entries form a Rademacher (or Steinhaus) sequence. Choose $b \in \mathbb{R}^N$ to be a Rademacher sequence. Let $y = \Phi^\Theta(b)x \in \mathbb{C}^m$. If

$$m \geq 57s \ln^2(17N^2/\varepsilon)$$

then $x$ can be recovered from $y$ via $\ell_1$-minimization with probability at least $1 - \varepsilon$. 
RIP estimate for partial random circulant matrices

**Theorem (Krahmer, Mendelson, Rauhut 2012)**

Let $\Theta \subset [N]$ be an arbitrary (deterministic) set of cardinality $m$. Choose $b \in \mathbb{R}^N$ to be a Rademacher sequence. Assume that

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

Then with probability at least $1 - N^{-\ln^2(s) \ln(N)}$ the restricted isometry constants of $\frac{1}{\sqrt{m}} \Phi^{\Theta}(b)$ satisfy $\delta_s \leq \delta$.

Previous bounds:
- Haupt, Bajwa, Raz (2008): $m \geq C\delta s^2 \ln N$.
- Rauhut, Romberg, Tropp (2010): $m \geq C\delta s^{3/2} \ln^{3/2}(N)$.
- Random sets $\Theta$, Romberg (2009): $m \geq C\delta s \ln^6 N$.

Result extends to (sub-)gaussian generators.
Translation and modulation on $\mathbb{C}^m$:

$$(T_kg)_j = g_{j-k \mod m} \quad (M_\ell g)_j = e^{2\pi i \ell j/m} g_j$$

Gabor synthesis matrix

$$A = (T_k M_\ell g)_{(k,\ell) \in [m]^2} \in \mathbb{C}^{m \times m^2}$$
Translation and modulation on $\mathbb{C}^m$:

$$(T_k g)_j = g_{j - k \mod m} \quad (M_\ell g)_j = e^{2\pi i \ell j / m} g_j$$

Gabor synthesis matrix

$$A = (T_k M_\ell g)_{(k, \ell) \in [m]^2} \in \mathbb{C}^{m \times m^2}$$

For subgaussian random vector $g$ the matrix $A_g$ satisfies RIP with high probability if

$$m \geq Cs \log^2 s \log^2 m$$

Analysis uses same tool as for partial random circulant matrices
Motivation I: Radar with moving objects

Received signal is superposition of delayed and Doppler shifted (i.e., translated and modulated) versions of sent signal.

Task: Determine distances and velocities from receive signal!
Recall that $\delta_s$ is smallest constant such that

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

for all $s$-sparse $x$.

Equivalently, with $T_s = \{x \in \mathbb{C}^N : \|x\|_2 \leq 1, \|x\|_0 \leq s\}$

$$\delta_s = \sup_{x \in T_s} \left| \|Ax\|_2^2 - \|x\|_2^2 \right|.$$
Recall that $\delta_s$ is smallest constant such that

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

for all $s$-sparse $x$.

Equivalently, with $T_s = \{x \in \mathbb{C}^N : \|x\|_2 \leq 1, \|x\|_0 \leq s\}$

$$\delta_s = \sup_{x \in T_s} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| .$$

For partial random circulant matrices,

$$Ax = \frac{1}{\sqrt{m}} R_\Omega(\epsilon * x) = \frac{1}{\sqrt{m}} R_\Omega(x * \epsilon) =: V_x \epsilon$$

with appropriate $V_x \in \mathbb{R}^{m \times N}$. Furthermore, $E\|V_x \epsilon\|_2^2 = \|x\|_2^2$.

Therefore, $\delta_s$ is supremum of a chaos process,

$$\delta_s = \sup_{x \in T_s} \left| \|V_x \epsilon\|_2^2 - E\|V_x \epsilon\|_2^2 \right| .$$
Let $\|B\|_F = \sqrt{\text{Tr}(B^*B)}$ denotes the Frobenius norm. For a set $\mathcal{B}$ of matrices, let

$$d_F(\mathcal{B}) = \sup_{B \in \mathcal{B}} \|B\|_F, \quad d_{2\rightarrow2}(\mathcal{B}) = \sup_{B \in \mathcal{B}} \|B\|_{2\rightarrow2},$$

$$I(\mathcal{B}, \|\cdot\|_{2\rightarrow2}) = \int_0^{d_{2\rightarrow2}(\mathcal{B})} \sqrt{\ln N(\mathcal{B}, \|\cdot\|_{2\rightarrow2}, t)} \, dt,$$

where $N(\mathcal{B}, \|\cdot\|_{2\rightarrow2}, t)$ denote the covering numbers.
Let \( \| B \|_F = \sqrt{\text{Tr}(B^*B)} \) denotes the Frobenius norm.
For a set \( B \) of matrices, let
\[
d_F(B) = \sup_{B \in B} \| B \|_F, \quad d_{2 \to 2}(B) = \sup_{B \in B} \| B \|_{2 \to 2},
\]
\[
\mathcal{I}(B, \| \cdot \|_{2 \to 2}) = \int_0 ^ {d_{2 \to 2}(B)} \sqrt{\ln N(B, \| \cdot \|_{2 \to 2}, t)} \, dt,
\]
where \( N(B, \| \cdot \|_{2 \to 2}, t) \) denote the covering numbers.

**Theorem (Krahmer, Mendelson, Rauhut 2012)**

Let \( B = -B \subset \mathbb{C}^{m \times N} \) be a symmetric set of matrices and \( \epsilon \) a Rademacher vector of length \( N \). Then
\[
\mathbb{E} \sup_{B \in B} \left| \| B \epsilon \|_2^2 - \mathbb{E} \| B \epsilon \|_2^2 \right| 
\leq C_1 \mathcal{I}(B, \| \cdot \|_{2 \to 2})^2 + C_2 d_{\| \cdot \|_F(B)} \mathcal{I}(B, \| \cdot \|_{2 \to 2}).
\]
Theorem (Krahmer, Mendelson, Rauhut 2012)

Let $\mathcal{B} \subset \mathbb{C}^{m \times N}$ with $\mathcal{B} = -\mathcal{B}$ and $\epsilon$ be a Rademacher vector. Then

$$
\mathbb{P} \left( \sup_{\mathcal{B} \in \mathcal{B}} \left| \| B \epsilon \|_2^2 - \mathbb{E} \| B \epsilon \|_2^2 \right| \geq C_1 E + t \right) \leq 2 \exp \left( -C_2 \min \left\{ \frac{t^2}{V^2}, \frac{t}{U} \right\} \right),
$$

where

$$
E := d_{\| \cdot \|_F (\mathcal{B}) \mathcal{I} (\mathcal{B}, \| \cdot \|_{2 \rightarrow 2}) + \mathcal{I} (\mathcal{B}, \| \cdot \|_{2 \rightarrow 2})^2}
$$

$$
V := d_{\| \cdot \|_{2 \rightarrow 2} (\mathcal{B}) (\mathcal{I} (\mathcal{B}, \| \cdot \|_{2 \rightarrow 2}) + d_{\| \cdot \|_F (\mathcal{B})}),
$$

$$
U := d_{\| \cdot \|_{2 \rightarrow 2} (\mathcal{B})}^2.
$$

Symmetry assumption $\mathcal{B} = -\mathcal{B}$ only made for convenience. Generalizes to random vectors of independent mean-zero subgaussian random variables.