A Tutorial on Recovery Conditions for Compressive System Identification of Sparse Channels

Borhan M. Sanandaji

Colorado School of Mines
University of California, Berkeley

Joint work with Tyrone Vincent, Michael Wakin, and Kameshwar Poolla
Systems with High-Dimensional but Sparse Representation

- Some systems are high-dimensional but have sparse representation
- Multipath propagation
- Sparse impulse response

http://www.kn-s.dlr.de
from input-output measurements \((a \text{ and } b)\), identify system \(x\)
from input-output measurements \((a \text{ and } b)\), identify system \(x\)

\[
\begin{align*}
    a &\quad \rightarrow \quad \text{System } x \\
    e &\quad \downarrow \\
    b &= a \ast x + e
\end{align*}
\]

A Priori Knowledge

The impulse response \(x \in \mathbb{R}^N\) is high-dimensional with few non-zero entries \(S\), where \(S \ll N\).
from input-output measurements \((a \text{ and } b)\), identify system \(x\)

\[
a \rightarrow \text{System } x \rightarrow b = a \ast x + e
\]

**A Priori Knowledge**

The impulse response \(x \in \mathbb{R}^N\) is high-dimensional with few non-zero entries \(S\), where \(S \ll N\).

**Goal**

Identify \(x\) from few measurements \(\sim b\) as short as possible.
Convolution Implies a Toeplitz Measurement Matrix

- each observation $b_i$ can be written as $b_i = \sum_{j=1}^{N} a_j x_{i-j}$
- collect $M$ measurements, put in a matrix-vector multiplication format

$$b = A x$$

- $x \in \mathbb{R}^N$ is $S$-sparse, it has $S \ll N$ non-zero entries
CS Exploits Sparsity!

**\( \ell_1 \)-minimization**

\[
\hat{x} = \arg \min_x \|x\|_1 \quad \text{subject to} \quad b = Ax
\]

where \( \|x\|_1 = \sum_i |x_i| \).

- convex! recovery via linear programming
- under which conditions on \( A \) the solution \( \hat{x} \) is the correct solution?
The Restricted Isometry Property (RIP)

Definition

A matrix $A \in \mathbb{R}^{M \times N}$ is said to satisfy the RIP of order $S$ with isometry constant $\delta_S \in (0, 1)$ if

$$(1 - \delta_S) \| \mathbf{x} \|_2^2 \leq \| A \mathbf{x} \|_2^2 \leq (1 + \delta_S) \| \mathbf{x} \|_2^2$$

holds for all $S$-sparse signals $\mathbf{x} \in \mathbb{R}^N$.

- sufficient condition for recovery

Recovery based on the RIP: Candès

If $A$ satisfies the RIP of order $2S$ with isometry constant $\delta_{2S} < \sqrt{2} - 1$, then it is possible to uniquely recover any $S$-sparse signal $\mathbf{x}$ from the measurements $\mathbf{b}$ solving an $\ell_1$-minimization problem.
Toeplitz matrices appear in our problems of interest

- **Challenge**: the sensing matrix $A$ now has a Toeplitz structure
- establish the RIP for a Toeplitz $A$
Outline

1. Concentration of Measure Inequalities

2. RIP based in Geršgorin Disk Theorem
Concentration of Measure Inequality (CoM)

- a simple way of proving the RIP for a randomized construction of $A$

**Definition**

A random matrix $A \in \mathbb{R}^{M \times N}$ is said to satisfy the CoM inequality if for any fixed signal $x \in \mathbb{R}^N$ (not necessarily sparse) and any $\epsilon \in (0, 1)$,

$$P\left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2 \right\} \leq 2e^{-Mf(\epsilon)},$$

where $f(\epsilon)$ is a positive constant that depends on $\epsilon$. 
Concentration of Measure Inequality (CoM)

- a simple way of proving the RIP for a randomized construction of $A$

**Definition**

A random matrix $A \in \mathbb{R}^{M \times N}$ is said to satisfy the CoM inequality if for any fixed signal $x \in \mathbb{R}^N$ (not necessarily sparse) and any $\epsilon \in (0, 1)$,

$$P \left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2 \right\} \leq 2e^{-Mf(\epsilon)},$$

where $f(\epsilon)$ is a positive constant that depends on $\epsilon$.

**CoM: Unstructured Gaussian Matrices**

If $A$ is populated with i.i.d. Gaussian entries with mean 0 and variance $\frac{1}{M}$, then $A$ satisfies the CoM inequality with $f(\epsilon) = \frac{\epsilon^2}{4}$. 
RIP based on CoM: Davenport

Let $\delta_S \in (0, 1)$ denote a distortion factor and $\nu \in (0, 1)$ denote a failure probability. Suppose $A$ satisfies the CoM inequality with

$$M \geq S \left( \log \left( \frac{42}{\delta_S} \right) + 1 + \log \left( \frac{N}{S} \right) \right) + \log \left( \frac{2}{\nu} \right) f \left( \frac{\delta_S}{\sqrt{2}} \right).$$

Then with probability at least $1 - \nu$,

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

holds for all $S$-sparse $x \in \mathbb{R}^N$. 
RIP based on CoM: Davenport

Let \( \delta_S \in (0, 1) \) denote a distortion factor and \( \nu \in (0, 1) \) denote a failure probability. Suppose \( A \) satisfies the CoM inequality with

\[
M \geq S \left( \log\left(\frac{42}{\delta_S}\right) + 1 + \log\left(\frac{N}{S}\right) \right) + \log\left(\frac{2}{\nu}\right) f\left(\frac{\delta_S}{\sqrt{2}}\right).
\]

Then with probability at least \( 1 - \nu \),

\[
(1 - \delta_S) \| x \|_2^2 \leq \| Ax \|_2^2 \leq (1 + \delta_S) \| x \|_2^2
\]

holds for all \( S \)-sparse \( x \in \mathbb{R}^N \).

RIP: Unstructured Gaussian Matrices

If \( A \) is populated with i.i.d. Gaussian entries with mean 0 and variance \( \frac{1}{M} \), then \( A \) satisfies the RIP of order \( S \) with isometry constant \( \delta_S \in (0, 1) \) with high probability when \( M \gtrsim \delta_S^{-2} S \log\left(\frac{N}{S}\right) \).
Theorem: S., Vincent, Wakin

Suppose \( A \in \mathbb{R}^{M \times N} \) is a Toeplitz matrix populated with i.i.d. Gaussian entries with mean 0 and variance \( \frac{1}{M} \). Then, for any fixed \( x \in \mathbb{R}^N \) and any \( \epsilon \in (0, 1) \),

\[
P \left\{ \left| \| Ax \|_2^2 - \| x \|_2^2 \right| \geq \epsilon \| x \|_2^2 \right\} \leq 2e^{-\frac{\epsilon^2 M}{8\rho(x)}},
\]

where \( \rho(x) := \max_i \frac{\lambda_i(P)}{\| x \|_2^2} \), is the normalized maximum eigenvalue of the covariance matrix of \( Ax, P(x) \).
- CoM for **Toeplitz** Gaussian Matrices:

\[
P \left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \geq \epsilon \|x\|_2^2 \right\} \leq 2e^{-\frac{\epsilon^2 M}{8\rho(x)}}
\]

- CoM for **unstructured** Gaussian Matrices:

\[
P \left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \geq \epsilon \|x\|_2^2 \right\} \leq 2e^{-\frac{\epsilon^2 M}{4}}
\]

- concentration exponent is worse by a factor of \(2\rho(x)\)
- the bounds are signal-dependent
  - signals with *scattered random sign* non-zero entries have small \(\rho(x)\)
  - signals with *clustered same sign* non-zero entries have large \(\rho(x)\)
Signal-Dependent Concentrations and Bounds

- fix two particular signals with $N = 1024$, $S = 64$
- generate 1000 i.i.d. Gaussian Toeplitz $A$ with $M = 512$
- $P \left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \geq \epsilon \|x\|_2^2 \right\}$

Clustered Same Sign

Scattered Random Sign
for all $S$-sparse signals $x \in \mathbb{R}^N$, $\rho(x) \leq S$

Corollary

Suppose $A \in \mathbb{R}^{M \times N}$ is a Toeplitz matrix populated with i.i.d. Gaussian entries with mean 0 and variance $\frac{1}{M}$. Then, for any $S$-sparse $x \in \mathbb{R}^N$ and any $\epsilon \in (0, 1)$,

$$
\mathbb{P} \left\{ \left\| Ax \right\|_2^2 - \| x \|_2^2 \geq \epsilon \| x \|_2^2 \right\} \leq 2e^{-\frac{\epsilon^2 M}{8S}} .
$$
RIP for Toeplitz Gaussian Matrices based on CoM

- for all $S$-sparse signals $x \in \mathbb{R}^N$, $\rho(x) \leq S$

**Corollary**

Suppose $A \in \mathbb{R}^{M \times N}$ is a Toeplitz matrix populated with i.i.d. Gaussian entries with mean 0 and variance $\frac{1}{M}$. Then, for any $S$-sparse $x \in \mathbb{R}^N$ and any $\epsilon \in (0, 1)$,

$$P \left\{ \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \geq \epsilon \|x\|_2^2 \right\} \leq 2e^{-\frac{\epsilon^2 M}{8S}}.$$

**RIP: Toeplitz Gaussian Matrices**

If $A$ is populated with i.i.d. Gaussian entries with mean 0 and variance $\frac{1}{M}$, then $A$ satisfies the RIP of order $S$ with isometry constant $\delta_S \in (0, 1)$ with high probability when $M \gtrsim \delta_S^{-2} S^2 \log\left(\frac{N}{S}\right)$. 
Recovery Performance and $\rho(x)$

- generate 1000 i.i.d. $M \times N$ ($N = 512$) Gaussian Toeplitz $A$
- find $M$ with over 99 percent recovery

- relation between $\rho(x)$ and the recovery performance
RIP for Toeplitz/Circulant Matrices

- Bajwa et al. - select enough samples so that i.i.d. rows exist:
  \[ M = O(S^3 \log\left(\frac{N}{S}\right)) \]

- Haupt et al. - analysis using Geršgorin disk theorem:
  \[ M = O(S^2 \log\left(\frac{N}{S}\right)) \]

- Rauhut et al. - Dudley’s inequality for chaos and generic chaining:
  \[ M = O(S^{1.5} \log(N)^{1.5}) \]

- Krahmer et al. - Dudley’s inequality for chaos and generic chaining:
  \[ M = O(S \log(S)^2 \log(N)^2) \]
Outline

1. Concentration of Measure Inequalities

2. RIP based in Geršgorin Disk Theorem
Definition

A matrix $A \in \mathbb{R}^{M \times N}$ is said to satisfy the RIP of order $S$ with isometry constant $\delta_S \in (0, 1)$ if

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

holds for all $S$-sparse signals $x \in \mathbb{R}^N$.

A Useful Interpretation

Establishing the RIP is equivalent to restricting all of the eigenvalues of all $\binom{N}{S}$ submatrices $A_S^T A_S \in \mathbb{R}^{S \times S}$ to the interval

$$(1 - \delta_S, 1 + \delta_S),$$

where $A_S \in \mathbb{R}^{M \times S}$ is a submatrix of $A \in \mathbb{R}^{M \times N}$ whose columns are those columns of $A$ indexed by the set $S$ with $|S| = S$. 
Geršgorin Disk Theorem

Consider a matrix $P \in \mathbb{R}^{N \times N}$. Let $r_i = \sum_{j=1}^{N} |P_{ij}|$. Then,

$$
\lambda_i(P) \in \bigcup_{i=1}^{N} D(P_{ii}, r_i), \quad i = 1, 2, \ldots, N,
$$

where $D(P_{ii}, r_i)$ is a disc centered at $P_{ii}$ with radius $r_i$. 
Sketch of the Proof

- define \( P = A^T A \in \mathbb{R}^{N \times N} \)
- assume \( |P_{ii} - 1| < \delta_1 \) for all \( N \) diagonal entries \( P_{ii} \)
- assume \( |P_{ij}| < \frac{\delta_2}{S} \) for all the \( \frac{N(N-1)}{2} \) distinct off-diagonal entries \( P_{ij} \)
- all eigenvalues of any \( S \times S \) submatrix of \( P \) formed by \( A_S^T A_S \) lie within
  \[
  \left( 1 - \delta_1 - (S - 1) \frac{\delta_2}{S}, 1 + \delta_1 + (S - 1) \frac{\delta_2}{S} \right)
  \]
- let \( \delta_1 + \delta_2 = \delta_S \), then all eigenvalues of all submatrices \( A_S^T A_S \) lie within
  \[
  (1 - \delta_S, 1 + \delta_S)
  \]
- derive tail probability bounds on \( P_{ii} \) and \( P_{ij} \)
Sketch of the Proof

- tail probability bounds
  - $\mathbb{P}\{|P_{ii} - 1| \geq \delta_1\} \leq 2e^{-\frac{\delta_1^2 M}{6}}$
  - $\mathbb{P}\{|P_{ij}| \geq \frac{\delta_2 S}{S}\} \leq 4e^{-\frac{\delta_2^2 M}{24S^2}}$
Sketch of the Proof

- Tail probability bounds
  - \( P \{|P_{ii} - 1| \geq \delta_1\} \leq 2e^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \{|P_{ij}| \geq \frac{\delta_2}{S}\} \leq 4e^{-\frac{\delta_2^2 M}{24S^2}} \)

- Union bound
  - \( P \left\{ \bigcup_{i=1}^{N} |P_{ii} - 1| \geq \delta_1 \right\} \leq 2Ne^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \left\{ \bigcup_{i=1}^{N} \bigcup_{j=1,j\neq i}^{N} |P_{ij}| \geq \frac{\delta_2}{S} \right\} \leq 4 \frac{N(N-1)}{2} e^{-\frac{\delta_2^2 M}{24S^2}} \leq 2N^2 e^{-\frac{\delta_2^2 M}{24S^2}} \)
Sketch of the Proof

- Tail probability bounds
  - \( P \{ |P_{ii} - 1| \geq \delta_1 \} \leq 2e^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \{ |P_{ij}| \geq \frac{\delta_2}{S} \} \leq 4e^{-\frac{\delta_2^2 M}{24S^2}} \)

- Union bound
  - \( P \left\{ \bigcup_{i=1}^{N} |P_{ii} - 1| \geq \delta_1 \right\} \leq 2Ne^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \left\{ \bigcup_{i=1}^{N} \bigcup_{j=1, j \neq i}^{N} |P_{ij}| \geq \frac{\delta_2}{S} \right\} \leq 4 \frac{N(N-1)}{2} e^{-\frac{\delta_2^2 M}{24S^2}} \leq 2N^2 e^{-\frac{\delta_2^2 M}{24S^2}} \)

- For some \( \delta_1 + \delta_2 = \delta_S \in (0, 1) \),

\[
P\{ A \text{ does NOT satisfy the RIP } (S, \delta_S) \} \leq 3N^2 e^{-\frac{\delta_S^2 M}{96S^2}}.
\]
Sketch of the Proof

- tail probability bounds
  - \( P \{|P_{ii} - 1| \geq \delta_1\} \leq 2e^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \{|P_{ij}| \geq \frac{\delta_2}{S}\} \leq 4e^{-\frac{\delta_2^2 M}{24S^2}} \)

- union bound
  - \( P \left\{ \bigcup_{i=1}^{N} |P_{ii} - 1| \geq \delta_1 \right\} \leq 2Ne^{-\frac{\delta_1^2 M}{6}} \)
  - \( P \left\{ \bigcup_{i=1}^{N} \bigcup_{j=1, j \neq i}^{N} |P_{ij}| \geq \frac{\delta_2}{S}\right\} \leq 4\frac{N(N-1)}{2}e^{-\frac{\delta_2^2 M}{24S^2}} \leq 2N^2e^{-\frac{\delta_2^2 M}{24S^2}} \)

- for some \( \delta_1 + \delta_2 = \delta_S \in (0, 1) \),

\[
P \{A \text{ does NOT satisfy the RIP} (S, \delta_S)\} \leq 3N^2e^{-\frac{\delta_S^2 M}{96S^2}}.
\]

- for a given \( \nu \in (0, 1) \), whenever

\[
M \geq 192\delta_S^{-2}S^2 \log \frac{\sqrt{3N}}{\nu},
\]

\[
P \{A \text{ satisfies the RIP} (S, \delta_S)\} \geq 1 - \nu^2
\]
Conclusion

- sparsity in systems and CSI
- system applications introduce structure
- establishing the RIP for Toeplitz Gaussian matrices
  - CoM inequality
  - Geršgorin disk theorem
- more details on CSI at http://inside.mines.edu/~bmolazem/

Thanks for your attention!
Sparse Signal Recovery

\( \ell_2 \)-Recovery

- \( \ell_2 \)-recovery (Euclidian distance) doesn’t work!

\[
\hat{x} = \arg\min_x \|x\|_2 \quad \text{s.t.} \quad b = Ax
\]

- minimum is almost never sparse

\[
\hat{x} = A'(AA')^{-1}b
\]
Sparse Signal Recovery

$\ell_2$-Recovery Geometry

$\mathbb{R}^N$

$x^*$

$\{x: Ax = b\}$
Sparse Signal Recovery

$\ell_2$-Recovery Geometry

Incorrect Recovery
Sparse Signal Recovery

ℓ₁ Recovery Geometry

\[ \{ x : Ax = b \} \]

\[ \{ x : \| x \|_1 \leq \| x^* \|_1 \} \]

Correct Recovery

\[ \{ x : Ax = b \} \]

\[ \{ x : \| x \|_1 \leq \| x^* \|_1 \} \]
Sparse Signal Recovery

$l_1$ Recovery Geometry

$\{x : Ax = b\}$

$\{x : \|x\|_1 \leq \|x^*\|_1\}$

Correct Recovery

Incorrect Recovery