On construction and analysis of sparse random matrices and expander graphs with applications to compressed sensing

Bubacarr Bah

École Polytechnique Fédérale de Lausanne (EPFL)

SUPERVISOR: Jared Tanner (University of Oxford, UK)

SampTA
Bremen, July 01, 2013
Compressed Sensing

- Signal $x \in \mathbb{R}^N$, $k$-sparse (sparse representation).
- Sensing matrix $A \in \mathbb{R}^{n \times N}$; measurements $y = Ax$, ($n \ll N$).
- Problem/solution: $\min_{x \in \mathbb{R}^N} \|x\|_0$ s.t. $Ax = y$. 

Encoder, $A$: Deterministic/probabilistic; dense/sparse?
Decoder, $\Delta$: $l_q$-minimizations & Greedy (SP, OMP, IHT, ...)

CS Applications
- Medical Imaging: MRI, Tomography, Radiology, ...
- Infrared spectroscopy & Seismic imaging
- Single pixel camera & Analog-to-digital converters
- DNA micro-arrays, radar, wireless communications, ...
Compressed Sensing

- Signal \( x \in \mathbb{R}^N, \ k\)-sparse (sparse representation).
- Sensing matrix \( A \in \mathbb{R}^{n \times N} \); measurements \( y = Ax, (n \ll N) \).
- Problem/solution: \( \min_{x \in \mathbb{R}^N} \|x\|_0 \) s.t. \( Ax = y \).
- Encoder, \( A \): Deterministic/probabilistic; dense/sparse?
- Decoder, \( \Delta \): \( l_q \)-minimizations & Greedy (SP, OMP, IHT, ...)

**Compressed Sensing (CS)**

- Restricted Isometry Property - RIP-\( p \)
- RIP-1 and Expander graphs

**Introduction**

- Construction of sparse random matrices
- Numerical results
- Conclusions

**Compressed Sensing Applications**

- Medical Imaging: MRI, Tomography, Radiology, ...
- Infrared spectroscopy & Seismic imaging
- Single pixel camera & Analog-to-digital converters
- DNA micro-arrays, radar, wireless communications, ...
Compressed Sensing

- Signal $x \in \mathbb{R}^N$, $k$-sparse (sparse representation).
- Sensing matrix $A \in \mathbb{R}^{n \times N}$; measurements $y = Ax$, $(n \ll N)$.
- Problem/solution: $\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{s.t.} \quad Ax = y$.
- Encoder, $A$: Deterministic/probabilistic; dense/sparse?
- Decoder, $\Delta$: $l_q$-minimizations & Greedy (SP, OMP, IHT, ...)

CS Applications

- Medical Imaging: MRI, Tomography, Radiology, ...
- Infrared spectroscopy & Seismic imaging
- Single pixel camera & Analog-to-digital converters
- DNA micro-arrays, radar, wireless communications, ...
Design and analysis of \((A, \Delta)\) with small \(n\) is key in CS, \(n = \mathcal{O}(k \log (N/k))\) is optimal.
Design and analysis of \((A, \Delta)\) with small \(n\) is key in CS, 
\(n = \mathcal{O}(k \log (N/k))\) is optimal.

**CS Tools of Analysis**

- Coherence  [Donoho & Huo; Elad & Bruckstein]
- Restricted isometry property (RIP)  [Candès & Tao]
- Nullspace property  [Donoho & Huo]
- Stochastic geometry  [Donoho; Donoho & Tanner]
• Design and analysis of \((A, \Delta)\) with small \(n\) is key in CS, 
\[ n = \mathcal{O}(k \log(N/k)) \] is optimal

**CS Tools of Analysis**

- Coherence [Donoho & Huo; Elad & Bruckstein]
- Restricted isometry property (RIP) [Candès & Tao]
- Nullspace property [Donoho & Huo]
- Stochastic geometry [Donoho; Donoho & Tanner]

**Definition (\(\ell_p\)-norm restricted isometry property)**

\[ A \] has RIP-\(p\) of order \(k\) if, for all \(k\)-sparse \(x\)

\[ (1 - L(k, n, N; A))\|x\|_p^p \leq \|Ax\|_p^p \leq (1 + U(k, n, N; A))\|x\|_p^p. \]

• RIP-\(p\) shows how close to isometry \(A\) is for \(k\)-sparse \(x\) in \(\| \cdot \|_p\)
RIP gives a sufficient guarantees for exact recovery, e.g. a subgaussian $A$ satisfies RIP-2 with $n = O(k \log (N/k))$ and if $R_k = \max [L(k, n, N; A), U(k, n, N; A)]$ using such $A$

$\ell_1$-minimization) works if:

- $R_{3k} + 3R_{4k} < 2$,  [Candès, Romberg & Tao, 2006]
- $R_{2k} < 2/(3 + \sqrt{7/4}) \approx 0.4627$,  [S. Foucart, 2010]
RIP gives a sufficient guarantees for exact recovery, e.g. a subgaussian $A$ satisfies RIP-2 with $n = O(k \log (N/k))$ and if $R_k = \max \{L(k, n, N; A), U(k, n, N; A)\}$ using such $A$.

\(\Delta (\ell_1\text{-minimization})\) works if:

1. $R_{3k} + 3R_{4k} < 2$, \cite{Candes-Romberg-Tao-2006}
2. $R_{2k} < 2/(3 + \sqrt{7/4}) \approx 0.4627$, \cite{Foucart-2010}

- A above is dense which is a computational bottleneck
- Sparse $A$: fast computation & low storage complexity
- Sparse $A$ do not satisfy RIP-2 with optimal $n$; e.g. for $A \in \{0, 1\}^{n \times N}$ we need $n = \Omega(k^2)$ \cite{Chandar-2007}
Sparse matrices of expander graphs satisfy RIP-1

**Theorem (RIP-1 - Berinde et. al., 2008)**

Consider $\Phi$ as the adjacency matrix of $(k, d, \epsilon)$-lossless expander such that $\{1/\epsilon, d\} < N$, then $A = d^{-1}\Phi$ satisfies RIP-1 with $L(k, n, N; A) = 2\epsilon$ and $U(k, n, N; A) = 0$
Sparse matrices of expander graphs satisfy RIP-1

Theorem (RIP-1 - Berinde et. al., 2008)

Consider $\Phi$ as the adjacency matrix of $(k, d, \epsilon)$-lossless expander such that $\{1/\epsilon, d\} < N$, then $A = d^{-1}\Phi$ satisfies RIP-1 with $L(k, n, N; A) = 2\epsilon$ and $U(k, n, N; A) = 0$
- Objects well-studied in theoretical computer science but there are construction issues

- **Probabilistic** construction with \( d = O \left( \log \left( \frac{N}{k} \right) / \varepsilon \right) \) and \( n = O \left( k \log \left( \frac{N}{k} \right) / \varepsilon^2 \right) \)

- **Explicit** construction with \( d = O \left( \left( \log N \right) \left( \log k \right) / \varepsilon \right)^{1+1/\alpha} \) and right set size \( \left( d^2 k^{1+\alpha} \right) \), for any \( \alpha > 0 \)  
  \[\text{[Guruswami et. al., 2007]}\]
• Objects well-studied in theoretical computer science but there are construction issues
  
• Probabilistic construction with \( d = \mathcal{O} \left( \log \left( \frac{N}{k} \right) / \epsilon \right) \) and \( n = \mathcal{O} \left( k \log \left( \frac{N}{k} \right) / \epsilon^2 \right) \)

• Explicit construction with \( d = \mathcal{O} \left( (\log N)(\log k) / \epsilon \right)^{1+1/\alpha} \) and right set size \( (d^2 k^{1+\alpha}) \), for any \( \alpha > 0 \) \[Guruswami et. al., 2007\]

**Goal of this work**

Probabilistically construct expander graphs using a novel dyadic splitting technique to compute better constants to be used in sampling theorems for

• existence of expander graphs and their adjacency matrices
• quantitative comparison of recovery algorithms
**Note** \( \text{Prob} \left( \|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1 \right) \equiv \text{Prob} \left( |A_k| \leq (1 - \epsilon)dk \right) \)
• **Note** $\text{Prob } (\|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1) \equiv \text{Prob } (|A_k| \leq (1 - \epsilon)dk)$

• **Dyadic splitting** of sets: $|A_k| = \left| A\left\lceil \frac{k}{2}\right\rceil \cup A\left\lfloor \frac{k}{2}\right\rfloor \right|$ and

$$\text{Prob } (|A_k| \leq (1 - \epsilon)dk) = \text{Prob } \left( \left| A\left\lceil \frac{k}{2}\right\rceil \cup A\left\lfloor \frac{k}{2}\right\rfloor \right| \leq (1 - \epsilon)dk \right)$$
Note \( \text{Prob} \left( \|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1 \right) \equiv \text{Prob} \left( |A_k| \leq (1 - \epsilon)dk \right) \)

Dyadic splitting of sets: \( |A_k| = \left| A\left\lceil \frac{k}{2} \right\rceil \cup A\left\lfloor \frac{k}{2} \right\rfloor \right| \) and

\[
\text{Prob} \left( |A_k| \leq (1 - \epsilon)dk \right) = \text{Prob} \left( \left| A\left\lceil \frac{k}{2} \right\rceil \cup A\left\lfloor \frac{k}{2} \right\rfloor \right| \leq (1 - \epsilon)dk \right)
\]

Given \( Z = X \cup Y \) s.t. \( |X| = x \) and \( |Y| = y \), then

(1) \( \text{Prob} \left( |X \cup Y| \leq z \right) = \sum_{l=\min(x,y)}^{\min(2\cdot\min(x,y),z)} \text{Prob} \left( |X \cup Y| = z \right) \)

(2) \( \text{Prob} \left( |X \cup Y| = z \right) = \text{Prob} \left( |X \cap Y| = x + y - z \right) \cdot \text{Prob} \left( |X| = x \right) \cdot \text{Prob} \left( |Y| = y \right) \)
Note $\text{Prob}\left(\|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1\right) \equiv \text{Prob}\left(|A_k| \leq (1 - \epsilon)dk\right)$

Dyadic splitting of sets: $|A_k| = \left|A\left\lceil \frac{k}{2} \right\rceil \cup A\left\lfloor \frac{k}{2} \right\rfloor\right|$ and

$$\text{Prob}\left(|A_k| \leq (1 - \epsilon)dk\right) = \text{Prob}\left(\left|A\left\lceil \frac{k}{2} \right\rceil \cup A\left\lfloor \frac{k}{2} \right\rfloor\right| \leq (1 - \epsilon)dk\right)$$

Given $Z = X \cup Y$ s.t. $|X| = x$ and $|Y| = y$, then

$$\text{(1)} \quad \text{Prob}\left(|X \cup Y| \leq z\right) = \sum_{l=\min(x,y)}^{\min(2\cdot\min(x,y),z)} \text{Prob}\left(|X \cup Y| = z\right)$$

$$\text{(2)} \quad \text{Prob}\left(|X \cup Y| = z\right) = \text{Prob}\left(|X \cap Y| = x + y - z\right) \cdot \text{Prob}\left(|X| = x\right) \cdot \text{Prob}\left(|Y| = y\right)$$

Next split $X$ and express $\text{Prob}\left(|X| = x\right)$ in terms of it’s children and their intersection, similarly $Y$
Repeat the splitting process until the last sets are made up of one column giving probability one since cardinalities are $d$.
Repeat the splitting process until the last sets are made up of one column giving probability one since cardinalities are $d$
End result: a product of nested sums of probabilities of the intersections which are computable
End result: a product of nested sums of probabilities of the intersections which are computable

Definition

Let $X \subset \Omega$ and $Y \subset \Omega$ s.t. $|X| = x$, $|Y| = y$ and $|\Omega| = n$, if $X$ and $Y$ are chosen uniformly at random independently then

$$\text{Prob} (|X \cap Y| = r) = \binom{x}{r} \binom{n-x}{y-r} \binom{n}{y}^{-1}$$
End result: a product of nested sums of probabilities of the intersections which are computable

**Definition**

Let \( X \subset \Omega \) and \( Y \subset \Omega \) s.t. \(|X| = x\), \(|Y| = y\) and \(|\Omega| = n\), if \( X \) and \( Y \) are chosen uniformly at random independently then

\[
\text{Prob}(|X \cap Y| = r) = \binom{x}{r}\binom{n - x}{y - r}\binom{n}{y}^{-1}
\]

- Need to **quantify** the sets and their cardinalities at each level using our **splitting lemma**
- Leading to the **tail bound** of \(|A_s|\)
Definition (Existence of expanders: $\rho^{\text{exp}}(\delta)$)

For $\epsilon$ and $d$ fixed, in the limit $(k, n, N) \to \infty$, $k/n \to \rho \in (0, 1)$ and $n/N \to \delta \in (0, 1)$, $\rho^{\text{exp}}(\delta)$ is the $\rho$ which makes

$$\text{Prob}\left( \|Ax\|_1 \geq (1 - 2\epsilon)d\|x\|_1 \right) \to 1$$

for all $k$–sparse $x$. 
Definition (Existence of expanders: $\rho^{\text{exp}}(\delta)$)

For $\epsilon$ and $d$ fixed, in the limit $(k, n, N) \to \infty$, $k/n \to \rho \in (0, 1)$ and $n/N \to \delta \in (0, 1)$, $\rho^{\text{exp}}(\delta)$ is the $\rho$ which makes $
abla \text{Prob}(\|Ax\|_1 \geq (1 - 2\epsilon)d\|x\|_1) \to 1$ for all $k$-sparse $x$.

- Approach is to upper bound $\text{Prob}(\|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1)$ and choose $\rho^{\text{exp}}(\delta)$ based on bound, i.e.

  $$\text{Prob}(\|Ax\|_1 \leq (1 - 2\epsilon)d\|x\|_1) \leq P(n; d, \epsilon) \cdot \exp [n \cdot \Psi(\delta, \rho; d, \epsilon)]$$

and define $\rho^{\text{exp}}(\delta)$ such that $\Psi(\delta, \rho; d, \epsilon) = 0$

- For $\rho < \rho^{\text{exp}}(\delta)$, $\Psi(\delta, \rho; d, \epsilon) < 0$ the probability that a randomly generated bipartite graph with these parameters is not an expander goes zero exponentially in $n$
Phase transitions: $\rho^\text{exp}(\delta)$

Varying $d$
Phase transitions: $\rho^{\text{exp}}(\delta)$

- Varying $d$
- Comparison to earlier bounds

$\rho^{\text{exp}}(\delta)$

$n = 1024, \epsilon = 0.25$

$d = 8$
$d = 12$
$d = 16$
$d = 20$
Comparisons of algorithms using performance guarantees

$\ell_1$ guarantees [Berinde et. al., 2008]

$\ell_1$ recovery is guaranteed if $\epsilon \leq 1/6$
Comparisons of algorithms using performance guarantees

$\ell_1$ guarantees [Berinde et. al., 2008]

$\ell_1$ recovery is guaranteed if $\epsilon \leq 1/6$
Comparisons of algorithms using performance guarantees

\( \ell_1 \) guarantees [Berinde et. al., 2008]

\( \ell_1 \) recovery is guaranteed if \( \epsilon \leq 1/6 \)
Comparisons of algorithms using performance guarantees

$\ell_1$ guarantees [Berinde et. al., 2008]

$\ell_1$ recovery is guaranteed if $\epsilon \leq 1/6$

$\ell_1$, SSMP, ER

Improved recovery more likely due to the closer match of the method of analysis

$\ell_1$: Gaussian vs. expander
Summary:

- Probabilistic construction using dyadic splitting resulting into better order constants
- Better sampling theorems (phase transitions) of expanders
- $\ell_1$ performance comparisons of algorithms
- Comparison of $\ell_1$ performance of Gaussian vs. expanders
- Improved recovery more likely related to method of analysis
Summary:
- Probabilistic construction using dyadic splitting resulting into better order constants
- Better sampling theorems (phase transitions) of expanders
- $\ell_1$ performance comparisons of algorithms
- Comparison of $\ell_1$ performance of Gaussian vs. expanders
- Improved recovery more likely related to method of analysis

References:
- B. Bah and J. Tanner. On construction and analysis of sparse random matrices and expander graphs with applications to compressed sensing. *SAMPTA 2013*
- B. Bah and J. Tanner. Vanishingly Sparse Matrices and Expander Graphs, with application to compressed sensing. Accepted into *IEEE IT, May 2013*
THANK YOU