Compressed Sensing with Sparse Matrices from Expander Graphs

Bubacarr Bah
with Prof. Jared Tanner

School of Mathematics, University of Edinburgh
and
Maxwell Institute

SIAM National Student Chapter Conference 2012 (SNSCC12)
Manchester
May 17, 2012
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. $Ax = y$.
- Algorithms: $l_q$ 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, ...
CS Tools of Analysis

- Coherence  [Donoho & Huo; Elad & Bruckstein]
- Restricted isometry property  [Candès & Tao]
- Nullspace property  [Donoho & Huo]
- Stochastic geometry  [Donoho; Donoho & Tanner]
- Message passing  [Donoho, Maleki & Montanari]

With the introduction of RIP$_1$, we refer to the standard RIP as RIP$_2$ - the subscripts 1 & 2 refer to the norms used

Definition (RIC$_2$)

RIC$_2$ of $A$ of order $k$ is the smallest number $R_k$, for all $k$-sparse $x$, such that

$$ (1 - R_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + R_k) \|x\|_2^2 $$
A having RIP<sub>2</sub> means that A is a near isometry for k-sparse x

RIP<sub>2</sub> gives a sufficient guarantees for **exact recovery**

**ℓ<sub>1</sub> minimization works if:**

- \( R_{3k} + 3R_{4k} < 2, \) \( [\text{Candès, Romberg & Tao, 2006}] \)
- \( R_{2k} < \sqrt{2} - 1, \) \( [\text{E. Candès, 2008}] \)
- \( R_{2k} < 2/(3 + \sqrt{7/4}) \approx 0.4627, \) \( [\text{S. Foucart, 2010}] \)

**Greedy Algorithms work if:**

- **IHT:** \( R_{3k} < 1/\sqrt{3}, \) \( [\text{S. Foucart, 2011}] \)
- **CoSaMP:** \( R_{4k} < \sqrt{2/(5 + \sqrt{73})}, \) \( [\text{S. Foucart, 2011}] \)
- **Subspace Pursuit (SP):** \( R_{3k} \lesssim 0.06, \) \( [\text{Dai & Milenkovic, 2009}] \)
A more quantitative definition is the asymmetric $\text{RIC}_2$:

**Definition (RIC$_2$)**

$\text{RIC}_2$ of $A$ of order $k$ is the smallest $L$ & $U$, for all $k$-sparse $x$, s.t.

\[
(1 - L(k, n, N; A))\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + U(k, n, N; A))\|x\|_2^2.
\]

- $\text{RIC}_2$ of $A$ & eigenvalues of $A^*_K A_K$, for $\Omega = \{1, 2, 3, \ldots, N\}$
  - \[
  1 + U(k, n, N; A) := \max_{K \subset \Omega, |K|=k} \lambda^{\max}(A^*_K A_K)
  \]
  - \[
  1 - L(k, n, N; A) := \min_{K \subset \Omega, |K|=k} \lambda^{\min}(A^*_K A_K)
  \]

- Thus $L$ & $U$ are smallest & largest deviation from unity of smallest & largest $\lambda(A^*_K A_K)$ respectively
Sparse matrices has fast and efficient implementation

Most sparse random matrices don’t satisfy RIP₂; e.g.: \{0, 1\} matrices do not satisfy RIP₂, unless \(n = \Omega(k^2)\) [Chandar, 07]

**Definition (RIP₁ - Berinde et. al., 2008)**

An \(n \times N\) matrix \(A\) satisfies RIP₁ if for any \(k\)-sparse vector \(x \in \mathbb{R}^N\) we have

\[
(1 - L(k, n, N; A)) ||x||_1 \leq ||Ax||_1 \leq ||x||_1
\]

**Theorem (RIP₁ - 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 RIP₁ with \(L(k, n, N; A) = C\epsilon\), where \(C > 1\)

Berinde et. al. showed that \(A\) of a \((k, d, \epsilon/2)\)-lossless expander satisfies RIP₁ with \(C = 2\)
Definition (Lossless Expander Graphs)

\[ G = (U, V, E) \] is an \((k, d, \epsilon)\)-lossless expander if it is a bipartite graph with \(|U| = N\) left vertices, \(|V| = n\) right vertices and has a regular left degree \(d\), such that any \(X \subset U\) with \(|X| \leq k\) has \(|\Gamma(X)| \geq (1 - \epsilon) d|X|\) neighbours.
Objects well-studied in theoretical computer science but there are construction issues

Probabilistic construction with $d = O\left(\log (N/k) / \epsilon\right)$ and $n = O\left(k \log (N/k) / \epsilon^2\right)$

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

Goal

Derive phase transitions, $\rho_{\text{exp}}(\delta)$, for the existence of lossless expander graphs and from $\rho_{\text{exp}}(\delta)$ derive $\rho_{\text{alg}}(\delta)$ such that in the limit of $n \to \infty$ with $\frac{n}{N} \to \delta \in (0, 1)$ and $\frac{k}{n} < (1 - \epsilon)\rho(\cdot)(\delta)$ it can be guaranteed that

- a randomly generated bipartite graph is an expander
- the output of an algorithm, $\hat{x}$, will satisfy $\|x - \hat{x}\|_1 \leq Const.\|x - x_k\|_1$
Linear growth or proportional-growth asymptotics

Problem instances \((k, n, N)\) considered is where the following ratios converge to nonzero bounded limits:

\[
\frac{k}{n} = \rho_n \to \rho \quad \text{and} \quad \frac{n}{N} = \delta_n \to \delta \quad \text{for} \quad (\delta, \rho) \in (0, 1)^2 \quad \text{as} \quad (k, n, N) \to \infty.
\]

Definition \((\rho^{\exp}(\delta))\)

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

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

\[
\Pr(\|Ax\|_1 \leq (1 - \epsilon)d\|x\|_1) \leq P(n; d, \epsilon) \cdot \exp [n \cdot \Psi(\delta, \rho; d, \epsilon)]
\]

and define \(\rho^{\exp}(\delta)\) such that \(\Psi(\delta, \rho; d, \epsilon) = 0\).
Bound derived from existing probabilistic construction give very small $\rho_{\text{exp}}(\delta)$; we improved this using **dyadic splitting of sets** since

$$\text{Prob}\left(\|Ax\|_1 \leq (1 - \epsilon)d\|x\|_1\right) = \text{Prob}\left(|A_k| \leq (1 - \epsilon)dk\right)$$

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

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

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

\begin{align*}
(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) \\
(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)
\end{align*}
Next we split $X$ and express $\text{Prob}(|X| = x)$ as a product of the probabilities of the cardinalities of its children and their intersection, similarly for $Y$.

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 computatable.

**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

$$\text{Prob}(|X \cap Y| = r) = \binom{x}{r}\binom{n-x}{y-r}\binom{n}{y}^{-1}$$
Combinatorial terms bounded using Stirling’s formula giving a bound of the probability as polynomial times exponential

Definition (Stirling’s inequality)

\[
\frac{16}{25} (2\pi p(1 - p)N)^{-\frac{1}{2}} e^{NH(p)} \leq \binom{N}{Np} \leq \frac{5}{4} (2\pi p(1 - p)N)^{-\frac{1}{2}} e^{NH(p)}
\]

where \( H(p) = -p \log(p) - (1 - p) \log(1 - p) \) be the Shannon entropy function for base \( e \) logarithms

The exponential is dominant and if the exponent is \( n \cdot \Psi(\cdot) \), the bound decays to zero exponentially in \( n \) when \( \cdot \Psi(\cdot) < 0 \), hence \( \rho^{\exp}(\delta) \) is defined s.t. \( \Psi(\delta, \rho; d, \epsilon) = 0 \)

For \( \rho < \rho^{\exp}(\delta) \), \( \Psi(\delta, \rho; d, \epsilon) < 0 \) the probability that a randomly generated bipartite graph with these parameters is not an expander goes zero exponentially in \( n \)
• Matches intuitive dependence on $\epsilon$ and $d$

\begin{align*}
\text{Dependence on } d & \\
\text{Dependence on } \epsilon &
\end{align*}
Comparisons to derivation using the other method

Other method

Dyadic set splitting
\( \ell_1 \) guarantees in sparse vs. dense

- \( \ell_1 \) recovery is guaranteed if \( \epsilon \leq 1/6 \) which is equivalent to \( L_k \leq 1/3 \) \([\text{Berinde et. al., 2008}]\)
- Phase transitions plots comparing \( \ell_1 \) performance

Other method

Dyadic set splitting
**Summary:**

- RIP$_1$ holds for (sparse) adjacency matrices of expanders which have preferable computation properties.
- Better phase transitions of existence of expander graphs using dyadic set splitting.
- $\ell_1$ performance guarantees of dense and sparse matrices.

**References:**

THANK YOU