On Model–Based RIP–1 Matrices

Piotr Indyk *Ilya Razenshteyn*

MIT, CSAIL

July 18, 2013

Compressive Sensing (Sparse Recovery)

- $x \in \mathbb{R}^n$ is a signal
- $Ax \in \mathbb{R}^m$ with $m \ll n$ is a linear sketch of x
- given Ax recover a good approximation to x provided that x is almost k–sparse (i.e., has at most k non-zero elements)
- formally, x^* s.t.

$$\|x - x^*\|_1 \le O(1) \cdot \min_{x' \text{ is } k\text{–sparse}} \|x - x'\|_1$$

- $m = ?$, $A = ?$
- motivation: signal acquisition, streaming and sketching algorithms, single-pixel cameras etc

RIP matrices

- A linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m \ll n$ such that for every k–sparse x $\|Ax\|_p \in (1 \pm \varepsilon)\|x\|_p$
- RIP options:
  - $p = 2$: random Gaussian with $m = O(k \log(n/k)/\varepsilon^2)$ rows (Candès, Romberg, Tao, 2006)
  - $p = 1$: (normalized) adjacency matrix of an unbalanced expander, $m = O(k \log(n/k)/\varepsilon^2)$ (Berinde, Gilbert, Indyk, Karloff, Strauss, 2008)
- These bounds are tight
- This implies sparse recovery with $m = O(k \log(n/k))$ measurements

Model–Based Compressive Sensing

- $x \in \mathbb{R}^n$ is a signal
- $Ax \in \mathbb{R}^m$ with $m \ll n$ is a linear sketch of x
- given Ax recover a good approximation to x provided that x is almost k–sparse with $\operatorname{supp} x$ having additional structure (Baraniuk, Cevher, Duarte, Hegde 2010)
- formally, x^* s.t.

$$\|x - x^*\|_1 \leq O(1) \cdot \min_{\substack{\operatorname{supp} x' \subseteq S \\ S \in \mathcal{M}_k}} \|x - x'\|_1,$$

  where $\mathcal{M}_k$ is a family of k–subsets of $[n]$
- $m = ?$, $A = ?$ (can potentially improve upon $m = O(k \log(n/k))$)

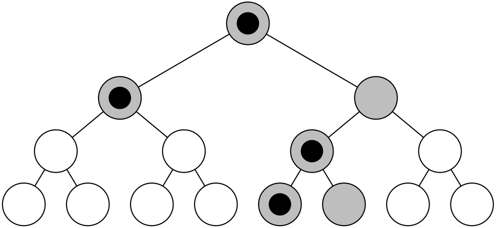
Model–based RIP matrices

- A model $\mathcal{M}_k$ is a family of k-subsets of $[n]$
- $A\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m \ll n$ such that for every x with $\operatorname{supp} x \subseteq S$ s.t. $S \in \mathcal{M}_k$ $\|Ax\| \in (1 \pm \varepsilon)\|x\|$
- If $\mathcal{M}_k$ is small, then can potentially have $m = o(k \log(n/k))$.

Two models: block sparsity

- n/b blocks of size b, signal occupies at most k/b of them
- represents "bursty" signals

Two models: tree sparsity



- signal is a subset of a connected subtree of size k
- represents images in the wavelet basis

Model–based RIP matrices for ℓ_2

- (Baraniuk, Cevher, Duarte, Hegde 2010): for any model $\mathcal{M}_k$ one can get a $\mathcal{M}_k$–RIP–2 matrix with $m = O((k + \log |\mathcal{M}_k|)/\varepsilon^2)$. In particular, $O(k)$ for block– and tree–sparse signals.
- The key idea: JL–type concentration inequality + union bound over ε–nets.
- (Indyk, Price 2011): can recover tree–sparse signals with $O(k)$ measurements and approximation $O(\sqrt{\log n})$. A similar result holds for block–sparse signals.

The main question

(Bertinoro list, 2011)

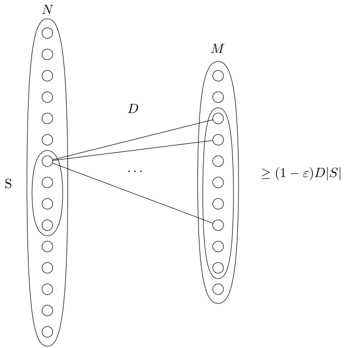
Can one achieve $O(k)$ measurements and $O(1)$ approximation for sparse recovery of tree– and block–sparse signals?

Our results

- Model–based RIP–1 matrices provide model–based sparse recovery with $O(1)$ approximation (with exponential time algorithm)
- There are block–sparse RIP–1 matrices with $m = O(k \log_k n)$ ($m = O(k)$ for $k = n^{\Omega(1)}$).
- There are tree–sparse RIP–1 matrices with $m = O(k \log(n/k) / \log\log(n/k))$
- Both these bounds are tight (for RIP matrices, not for sparse recovery in general!)

Upper bounds

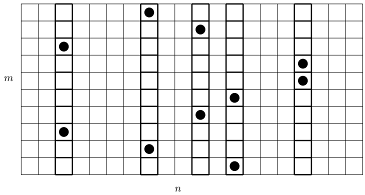
- $G = (U, V, E)$ is an unbalanced expander for a model $\mathcal{M}_k$, if U is d-regular and every subset S of an element of $\mathcal{M}_k$ has at least $(1-\varepsilon)d|S|$ neighbors
- Generalizes standard unbalanced expanders
- Normalized adjacency matrix of G has RIP–1 property for $\mathcal{M}_k$
- simple probabilistic arguments: $m = O(k \log_k n)$ for block–sparse and $m = O(k \log(n/k)/ \log\log(n/k))$ for tree–sparse.



Lower bounds

- Generalized expanders
- Sparsification: reduce the column sparsity to $O(m/k)$
- Counting argument

Generalized expanders



- $A \in \mathbb{R}^{m \times n}$ with each column having ℓ_1 norm at most 1 expands a set $S \subseteq [n]$, if

$$\sum_{i \in [m]} \max_{j \in S} |a_{ij}| \geq (1 - \varepsilon)|S|$$

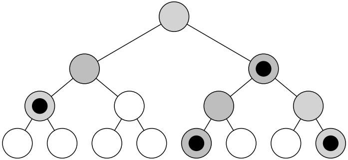
- Intuition: large entries in different columns barely conflict in rows
- If A is RIP–1 for a model $\mathcal{M}_k$, then it expands all sets $S \subseteq M$, $M \in \mathcal{M}_k$.

Sparsification

- For every tree–sparse RIP–1 matrix there is a constant fraction of columns s.t. one can zero out everything except at most $O(m/k)$ entries per column
- Proof idea: partition columns into n/k blocks of size k, observe that A expands these blocks, then for every block leave the largest entry in each row

Counting argument I

- Every tree–sparse RIP–1 matrix is RIP–1 for all sets of size $l = k / \log(n/k)$
- Column sparsity is $s = O(m/k)$
- Need to understand the tradeoff between m and s



Counting argument II

Following (Nachin 2010)
(see also (Do Ba, Price, Indyk, Woodruff 2010) and (Nelson, Nguyễn 2013))

- A is an almost isometry for l–sparse vectors and has column sparsity s
- $U \subset \mathbb{R}^n$ — a set of $l/2$-sparse vectors of ℓ_1–norm $O(1)$ and $\Omega(1)$ pairwise ℓ_1–distances, $|U| = \exp(l \log(n/l))$
- Existence of U is via error–correcting codes
- $T = \{Au\}_{u \in U}$
  - T is a set of $sl/2$-sparse vectors
  - T is a set of vectors of bounded length
  - Any two elements of T are at distance $\Omega(1)$

Counting argument III

- T is a set of vectors from $\mathbb{R}^m$ that are $sl/2$–sparse
- $|T| = \exp(l \log(n/l))$
- By the pigeonhole principle, $T' \subseteq T$ with $|T'| \geq |T| / \binom{m}{sl/2}$ such that all elements of T' have the same support
- Vectors of T' have bounded length and pairwise distances $\Omega(1)$
- Thus, $|T'| \leq \exp(sl)$ and since $s = O(m/k)$ and $l = k/\log(n/k)$

$$\begin{aligned} \exp(sl) &\geq |T'| \geq \frac{|T|}{\binom{m}{sl/2}} \geq \exp(l \log \frac{n}{l} - sl \log \frac{m}{sl}) \\ s &\geq \Omega\left(\frac{\log \frac{n}{l}}{\log \frac{m}{sl}}\right) \geq \Omega\left(\frac{\log(n/k)}{\log\log(n/k)}\right) \end{aligned}$$

Conclusion and open problems

- Our results
  - Model–based RIP–1 matrices allow model–based sparse recovery with $O(1)$ approximation
  - Tight bounds for block– and tree–sparse RIP–1 matrices (improving upon $O(k \log(n/k))$
- Open problems
  - Efficient recovery algorithm for model–based RIP–1 matrices
  - Settle #measurements for model–based compressive sensing

Questions?