The Restricted Isometry Property (RIP) of Subsampled Fourier Matrices

Ishay Haviv

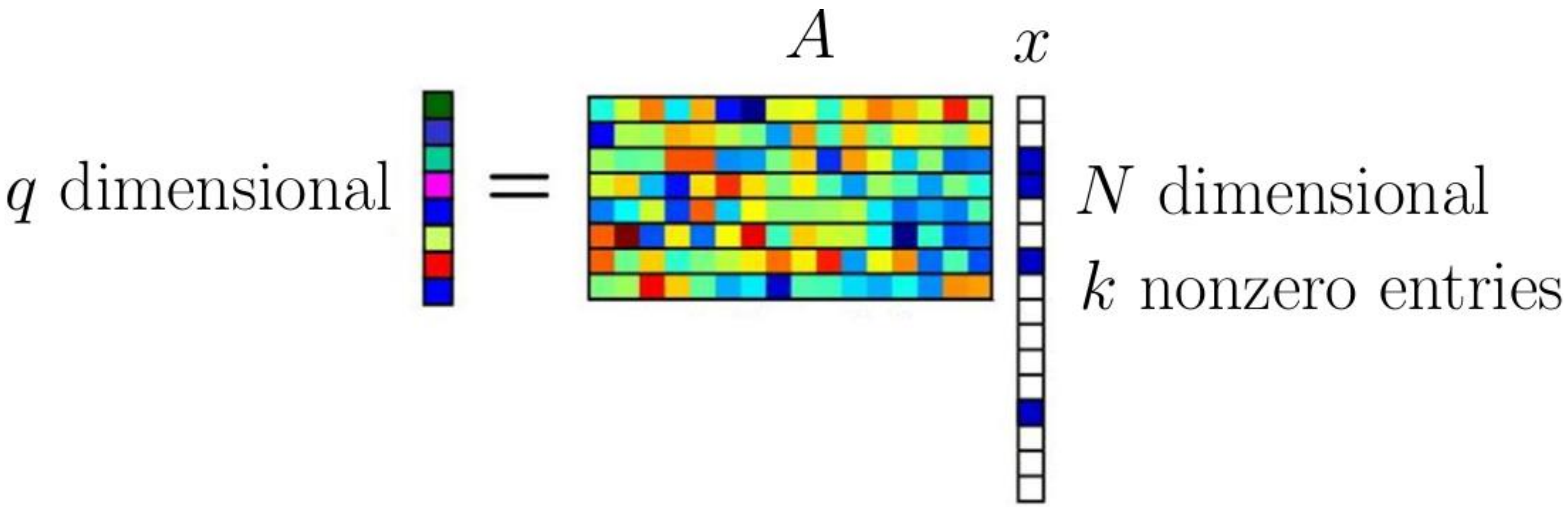
The Academic College of Tel Aviv-Yaffo

Oded Regev

Courant Institute
New York University

Sparse Recovery

- Let A be a q×N matrix (N>>q)
- Task: given Ax $\in \mathbb{R}^q$ for a high-dimensional *sparse* vector x $\in \mathbb{R}^N$ (i.e., at most k nonzero coordinates), recover x
- Many applications in TCS, applied math, signal processing: compressed sensing, streaming, coding theory, norm embeddings, computational complexity,…



The Restricted Isometry Property

- Def [CandésTao'06]: A q×N matrix A satisfies the *restricted isometry property of order k* if $||Ax||_2 \approx ||x||_2$ for all k-sparse vectors x
  - Essentially: any k columns of A are nearly orthogonal

- Thm [CandésTao'06]: The restricted isometry property of order 2k is a sufficient condition for sparse recovery
  - Moreover, recovery can be done *efficiently* using convex programming (and other more practical methods)

- Goal: minimize the number of rows q as a function of N and k
- Random matrices achieve the *optimal* number of rows

$$q = \Theta(k \cdot \log(N/k))$$

The Restricted Isometry Property of Fourier Matrices

- However, in many applications, A is a random sample of rows from the Discrete Fourier Transform matrix (MRI, ultrasound imaging, telescopes,…)
  - An advantage of using such an A is that we get very efficient compression and reconstruction algorithms using FFT
- Known upper bounds:
  - $q = O(k \cdot \log^6 N)$ [CandésTao'06]
  - $q = O(k \cdot \log^2 k \cdot \log(k \cdot \log N) \cdot \log N)$ [RudelsonVershynin'08]
  - $q = O(k \cdot \log^3 k \cdot \log N)$ [CheraghchiGuruswamiVelingker'13]
  - $q = O(k \cdot \log k \cdot \log^2 N)$ [Bourgain'14]
- Our result improves on all previous work:

$$q = O(k \cdot \log^2 k \cdot \log N)$$

– All these results apply to any orthogonal matrix with entries $O(N^{-1/2})$

Techniques

- Like all previous work on this question, our proof can be seen as a careful union bound applied to a sequence of progressively finer nets (i.e., chaining!)
- However, we avoid the use of Gaussian processes, the "symmetrization process," and Dudley's inequality as in most previous proofs
- Instead our proof is more in line with Bourgain's proof, and is (arguably) more elementary and short (~3.5 pages)

The proof

Proof Idea – The Setup

- M is an $N \times N$ orthogonal matrix with entries $O(1/\sqrt{N})$
- Let $Q \subseteq [N]$ be randomly chosen subset of q elements
- Let A be the rows of M specified by Q (scaled by $\sqrt{N/q}$)
- Goal: with high probability, for all k-sparse x,

$$\|Ax\|_2^2 \approx \|x\|_2^2.$$

- Equivalently,

$$\|Ax\|_2^2 \approx \|Mx\|_2^2.$$

- Yet another way of expressing this condition is as

$$\mathop{\mathbb{E}}_{j \in Q}\left[(Mx)_j^2\right] \approx \mathop{\mathbb{E}}_{j \in [N]}\left[(Mx)_j^2\right] ,$$

  i.e., that a sample $Q \subseteq [N]$ of q coordinates of the vector $(Mx)^2$ gives a good approximation to the average of all its coordinates.

Proof Idea – Very High Level

At a high level, the proof proceeds by:

1. Define a finite set of vectors $\mathcal{H}$ that forms a *net*, i.e., a set satisfying that any vector $(Mx)^2$ is close to one of the vectors in $\mathcal{H}$.
2. Argue using the Chernoff-Hoeffding bound that for any fixed vector $h \in \mathcal{H}$, a sample of q coordinates gives a good approximation to the average of h.
3. Union bound over all $h \in \mathcal{H}$.

Proof Idea – Some more details

Recall that our goal is:

$$\mathop{\mathbb{E}}_{j \in Q}\left[(Mx)_j^2\right] \approx \mathop{\mathbb{E}}_{j \in [N]}\left[(Mx)_j^2\right] . \tag{1}$$

- For convenience, assume w.l.o.g. that all entries of M are ± 1.
- For convenience, assume w.l.o.g. that $\|x\|_1 = 1$.
- Since x is k-sparse, $\|x\|_2^2$ is not too small ($\geq 1/k$), and therefore so is the right-hand side of (1).

To define $\mathcal{H}$:

1. Notice that since $\|x\|_1 = 1$, we can think of Mx as the *expectation* of a certain probability distribution over the columns of M (possibly with signs).
2. Using the Chernoff-Hoeffding bound again, this implies that we can approximate Mx well by taking the average over a small number of samples from this distribution. ("Maurey's empirical method")
3. We then let $\mathcal{H}$ be the set of all possible such averages, and obviously $|\mathcal{H}| \leq N^{\#\text{samples}}$.

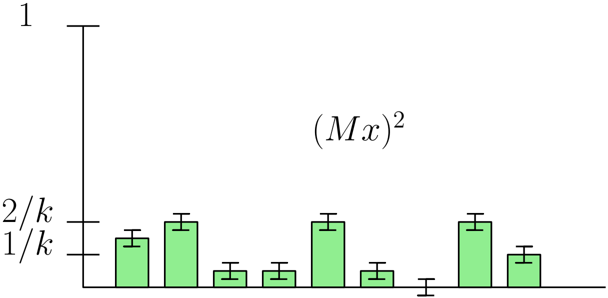
Proof Idea – the net

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0.1 \\ 0 \\ 0.7 \\ 0 \\ 0 \\ 0.2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.6 \\ -0.4 \\ -0.8 \\ 0.4 \\ 0.8 \\ -1 \\ -0.6 \end{pmatrix}$$

Proof Idea – the net

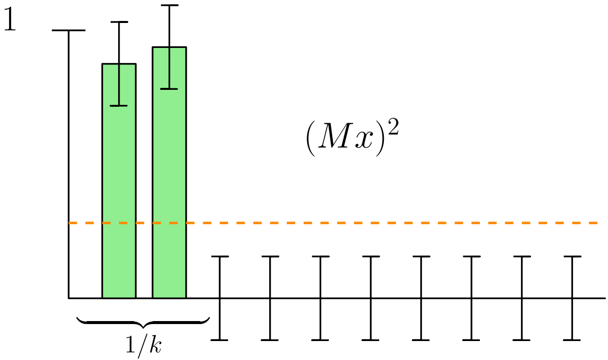
$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2/3 \\ 0 \\ 0 \\ 1/3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/3 \\ -1/3 \\ -1 \\ 1/3 \\ 1 \\ -1 \\ -1/3 \end{pmatrix}$$

Proof Idea – net for the flat case



- Here we need to approximate $(Mx)^2$ to within $\pm 1/k$
- For that, $\approx k$ samples suffice
- Therefore size of net is $\approx N^k$
- To properly sample from all vectors in net therefore requires $q \geq k \log N$ (only constantly many samples are required for each fixed vector, but we need confidence of N^{-k} for the union bound).☺

Proof Idea – net for the peaked case



- Here it suffices to approximate $(Mx)^2$ to within, say, $\pm 1/10$.
- However, noise will flood the result
- We therefore *threshold* the result and assume that coordinates below, say, $1/4$ are zero
- In order to have $\ll 1/k$ false positives, need $\approx \log k$ samples
- Therefore size of net is $\approx N^{\log k}$
- To sample from all vectors in net need $q \geq k \log k \log N$.☺

Proof Idea – summary

- We consider several nets $\mathcal{H}_1, \mathcal{H}_2, \ldots$, each responsible for approximating a different scale of $(Mx)^2$.
- For each x, we can approximately decompose $(Mx)^2$ as $\sum_i h^{(i)}$ with $h^{(i)} \in \mathcal{H}_i$
- We show that $q \geq k(\log k)^2 \log N$ suffices to approximate all vectors in all $\mathcal{H}_i$

Dependence on ε

- In most previous work, q has a ε^{-2} term
- In Bourgain's work, it was a much worse ε^{-6}
- The proof we just saw gives ε^{-4}
- By refining the proof we can recover the ε^{-2} dependence (up to logarithmic terms)

Conclusion

- We proved: q = O(k · log²k · logN) random rows of the Fourier matrix suffice for RIP of order k
- Best known lower bound Ω(k · logN) [BandeiraLewisMixon'15]
- Main open question: close the gap

- Another question: can our proof be cast in the Gaussian framework of Rudelson and Vershynin?

Thanks!