Randomized Algorithms 2015A
Lecture 12 – Compressed Sensing and RIP matrices*

Robert Krauthgamer

1 Compressed Sensing

Problem definition: We wish to learn an unknown vector $x \in \mathbb{R}^n$ through linear measurements, which means we choose a vector $a \in \mathbb{R}^n$ and observe the inner-product $a^T x$.

We want to minimize m, the number of linear measurements. If they are non-adaptive, then the measurement algorithm (without decoding part) can be described as a matrix $A \in \mathbb{R}^{m \times n}$.

Naive solution: Any choice of $m = n$ linear measurements that are linearly independent (i.e., A is invertible) is clearly sufficient (and also necessary).

Sparsity: We may know (by "prior information") that x is k-sparse, i.e., has at most k non-zeros. We will actually focus on almost k-sparse vector in the sense that $x = x' + z$ where x' is sparse and z is "noise", say $\|z\|_1$ is small. This is essentially a linear sketch for sparse inputs.

Exer: See if the results about sketching heavy hitters can be used here and what bounds do they imply.

Turns out that $m = O(k \log n)$ measurements suffice, and A can be taken to be a matrix of independent Gaussians.

Algorithmic approach: Recall we are given the vector of observations, which is the product Ax. Under exact k-sparsity $\|x\|_0 \le k$, an ideal algorithm could be to solve

$$\min\{\|x^*\|_0 : Ax^* = Ax\}.$$

In the general case, our algorithm will minimize instead the ℓ_1-norm

$$\min\{\|x^*\|_1 : Ax^* = Ax\}.$$

Exer: Verify that solving this problem (computing x^*) can be done in polynomial time using linear programming.

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

Theorem 1: Define $E_1^k(x) = \min\{\|x - x'\|_1 : x' \text{ is } k\text{-sparse}\}$. Then with probability at least $1 - 2/n$ over the choice of A,

$$\|x^* - x\|_2 = O(E_1^k(x)/\sqrt{k}). \tag{1}$$

This formalizes the scenario mentioned above, with $E_1^k(x) = \|z\|_1$, and then the approximation to true x' depends only on magnitude of noise. In particular, if x was exactly k-sparse, then we obtain exact recovery.

The statement provides a so-called "for each" guarantee – for each $x \in \mathbb{R}^n$, with high probability the approximation (1) holds. We will actually prove something stronger, called "for all" guarantee – with high probability the approximation (1) holds for all $x \in \mathbb{R}^n$.

RIP: The key will be to prove that WHP A has the following property: A matrix $A \in \mathbb{R}^{m \times n}$ satisfies the (k, δ)-Restricted Isometry Property (RIP) if for every k-sparse vector $x \in \mathbb{R}^n$,

$$(1 - \delta)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \delta)\|x\|_2. \tag{2}$$

Remark: this condition is equivalent to requiring that for every submatrix of A a consisting of k columns, all the singular values lie in the range $[1 - \delta, 1 + \delta]$.

The theorem follows immediately from the following two theorems.

Theorem 2: For suitable $m = O(k \log n)$, if the entries of $A \in \mathbb{R}^{m \times n}$ are independent Gaussians with distribution $N(0, 1/k)$, then with property at least $1 - 2/n$, matrix A is $(k, 1/3)$-RIP.

Theorem 3: If A is $(25k, 1/3)$-RIP then (for all x) (1) holds.

2 Constructing RIP matrix (Proof of Theorem 2)

Claim 4 (Crude Bound): WHP,

$$\forall x \in \mathbb{R}^n, \qquad \|Ax\|_2 \leq n^2\|x\|_2.$$

The proof, based on straightforward calculation, was seen in class.

Proof of theorem 2: The proof seen in class is based on a union bound over all subsets $T \subset [n]$ of size $|T| = k$; for each such T, the problem reduces to proving that for a matrix $B \in \mathbb{R}^{m \times k}$ of independent Gaussians $N(0, 1/k)$, with high probability

$$\forall y \in \mathbb{R}^k, \qquad \tfrac{2}{3}\|y\|_2 \leq \|By\|_2 \leq \tfrac{4}{3}\|y\|_2. \tag{3}$$

The latter is achieved by discretizing the unit sphere in $\mathbb{R}^k$, using Claim 5 below, applying on that discrete set the JL-lemma, and then extending the bound to the entire sphere. Overall, we get a failure probability $\binom{n}{k}|P|2^{-\Omega(m)} \leq 2^{O(k \log n) - \Omega(m)} \leq 1/n$, which proves Theorem 2.

Claim 5: For every $\varepsilon \in (0, 1)$ there is a set $P \subset S$ of size $O(1/\varepsilon)^k$ that is an ε-net of S, i.e., for every $x \in S$ there is $p \in P$ such that $\|p - x\| \leq \varepsilon$.

Exer: Does the analysis above actually work for $m = O(k \log \frac{n}{k})$? (This is effective to beat the trivial bound $m = n$ when k is "large".)

Exer: Let the matrix $A \in \mathbb{R}^{n \times n}$ have independent $\{\pm 1\}$ entries. Prove that with high probability $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$ is at most $O(\sqrt{n \log n})$. (Using one more idea, it is actually possible to prove a better bound of $O(\sqrt{n})$.)

3 ℓ_1-decoding (Proof of Theorem 3)

To simplify notation, let $h = x^* - x$, and recall our goal is to bound $\|h\|_2$. WLOG order the coordinates such that

- $|x_1|, \ldots, |x_k|$ are all at least $|x_{k+1}|, \ldots, |x_n|$.
- $|h_{k+1}| \geq \cdots \geq |h_n|$.

Define the sets of indices

- $T_0 = \{1, \ldots, k\}$
- $T_1 = \{k+1, \ldots, 26k\}$
- $T_2 = \{26k+1, \ldots, 51k\}$,

and so forth. Notice that $|T_0| = k$ and $|T_i| = 25k$ for all $i \geq 1$.

Define also $T_{01} = T_0 \cup T_1$, and $\overline{T_0} = [n] \setminus T_0$. Let X_T be the restriction of x to coordinates in the set T, and define (recall our ordering)

$$\varepsilon = E_1^k(x) = \|x_{\overline{T_0}}\|_1.$$

Recall that our goal is to bound $\|h\|_2 \leq O(1/\sqrt{k})\varepsilon$.

Claim 6: $\|h_{\overline{T_0}}\|_1 \leq \|h_{T_0}\|_1 + O(\varepsilon)$.

Claim 7: $\|h_{\overline{T_{01}}}\|_2 \leq \|h_{T_0}\|_2 + O(\varepsilon/\sqrt{k})$.

Claim 8: $\|h_{T_{01}}\|_2 \leq O(\varepsilon/\sqrt{k})$.

Proof of Theorem 3: Using triangle inequality, then Claim 7 and then 8,

$$\|x^* - x\|_2 = \|h\|_2 \leq \|h_{T_{01}}\|_2 + \|h_{\overline{T_{01}}}\|_2 \leq 2\|h_{T_{01}}\|_2 + O(\varepsilon/\sqrt{k}) \leq O(\varepsilon/\sqrt{k}).$$

QED.

We did not cover in class the proof of the three claims above; their proof can be found in Nick Harvey's lecture notes (Lecture 8). (Claim 6 is needed to prove of Claims 7 and 8.)