Sublinear Time and Space Algorithms 2018B – Lecture 6
Compressed Sensing via RIP matrices and Basis Pursuit*

Robert Krauthgamer

1 RIP matrices

Definition: A matrix $A \in \mathbb{R}^{m\times n}$ is (k,ε)-RIP (satisfies the restricted isometry property) if for every k-sparse vector $x \in \mathbb{R}^n$,

$$(1-\varepsilon)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\varepsilon)\|x\|_2^2.$$

Another interpretation: Let A_S denote the restriction of A to columns in $S \subset [n]$. Then the above requires that for all S of cardinality k, and all $x \in \mathbb{R}^S$, we have

$$(1-\varepsilon)\|x\|_2^2 \le xA_S^T A_S x \le (1+\varepsilon)\|x\|_2^2,$$

which means that $A_S^T A_S \approx I_k$ in the sense that all its eigenvalues are close to 1. We can further write it as $|x^T(A_S^T A_S - I)x| \le \varepsilon\|x\|_2^2$, which in matrix notation is just a bound on the operator norm (spectral radius):

$$\|A_S^T A_S - I\| \le \varepsilon.$$

Exer: Show that that this implies A_S is invertible.

Exer: Show that every (ε/k)-coherent matrix is (k,ε)-RIP.

Recall that a matrix $A \in \mathbb{R}^{m\times n}$ is called α-coherent if its columns A^i satisfy that every $\|A^i\|_2 = 1$ and every $|\langle A^i, A^j\rangle| \le \varepsilon$ (for $i \ne j$).

By the homework exercise, this implies that for every (n,k,ε), there exists a (k,ε)-RIP matrix with $m = O(\varepsilon^{-2}k^2 \log n)$ rows.

Hint: Given A that is α-coherent matrix for $\alpha = \varepsilon/k$, let $B = A_S^T A_S - I$, and bound $\|B\|$ which is the largest-magnitude eigenvalue of B.

*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.

2 Compressed Sensing via Basis Pursuit

Theorem 1 [Candes, Romberg and Tao [2004], and Donoho [2004]: There is a polynomial-time algorithm that given a matrix $A \in \mathbb{R}^{m \times n}$ which is $(2k, \varepsilon)$-RIP for $1 + \varepsilon < \sqrt{2}$, together with $y = Ax$ for some (unknown) $x \in \mathbb{R}^n$, computes $\tilde{x} \in \mathbb{R}^n$ satisfying

$$\|x - \tilde{x}\|_2 \leq O(1/\sqrt{k}) \|x_{tail(k)}\|_1.$$

This condition is usually called an ℓ_2/ℓ_1 guarantee.

Exer: Show that the above implies the following ℓ_1/ℓ_1 guarantee for $x^* = \tilde{x}_{top(k)}$:

$$\|x - x^*\|_1 \leq O(1) \|x_{tail(k)}\|_1.$$

Hint: Let T be the top k coordinates of x, and $\tilde{T}$ the top k coordinates of $\tilde{x}$. Split the coordinates into $\tilde{T}$, $T \setminus \tilde{T}$, and the rest.

Comparison with previously seen result: We saw previously an algorithm of [Cormode and Muthukrishnan, 2006] achieving WHP ℓ_1/ℓ_1 guarantee

$$\|x - x'\|_1 \leq (1 + 3\varepsilon) \|x_{tail(k)}\|_1.$$

* The current ℓ_2/ℓ_1 guarantee is stronger as it implies an ℓ_1/ℓ_1 guarantee, although with constant factor and not $1 + 3\varepsilon$.

* The current result is deterministic and holds for all x simultaneously, while the previous one holds WHP separately for every x.

* Previously, the number of measurements was $m = O(\varepsilon^{-1} k \log n)$. Here it depends on having an RIP matrix; the incoherent matrix from homework has worse (quadratic) dependence on k, but other constructions of RIP matrices are linear in k.

Basis Pursuit Algorithm: We will prove Theorem 1 using an algorithm called Basis Pursuit, which simply solves the linear program (LP)

$$\tilde{x} = \min\{\|z\|_1 : z \in \mathbb{R}^n, Az = y\}.$$

It is known that linear programs can be solved in polynomial time.

Exer: Show that $\tilde{x}$ above can indeed be solved using LP.

Proof of Theorem 1 (based on [Candes'08]):

As before, let z_S denote a vector z after zeroing all coordinates outside $S \subset [n]$.

Let $T_0 \subset [n]$ be the indices of the k heaviest coordinates (largest in absolute value) in x. Thus $x_{T_0^c} = x_{tail(k)}$.

We now partition the rest (T_0^c) according to the heaviness in $h = x - \tilde{x}$ (not in x). Let $T_1 \subset T_0^c$ be the k heaviest coordinates in $h_{T_0^c}$ largest ones (i.e., largest in T_0^c), and so forth.

To bound the error of $h = x - \tilde{x}$, we use the triangle inequality

$$\begin{aligned}\|x - \tilde{x}\|_2 = \|h\|_2 &= \|h_{T_0 \cup T_1} + h_{(T_0 \cup T_1)^c}\|_2 \\ &\le \|h_{T_0 \cup T_1}\|_2 + \|h_{(T_0 \cup T_1)^c}\|_2.\end{aligned}$$

The proof will be completed by the following two lemmas.

QED

Lemma 1a: $\|h_{(T_0 \cup T_1)^c}\|_2 \le O(1/\sqrt{k})\|x_{tail(k)}\|_1 + \|h_{T_0 \cup T_1}\|_2$.

Lemma 1b: $\|h_{T_0 \cup T_1}\|_2 \le O(1/\sqrt{k})\|x_{tail(k)}\|_1$.

We prove these two lemmas using ... another lemma.

Lemma 1c: $\sum_{j \ge 2} \|h_{T_j}\|_2 \le \frac{2}{\sqrt{k}} \cdot \|x_{T_0^c}\|_1 + \|h_{T_0 \cup T_1}\|_2$.

Proof of Lemma 1c: Was seen in class using the so-called "shelling argument" and the fact that Now $\tilde{x} = x - h$ is a minimizer of the LP, and x is feasible.

Proof of Lemma 1a: was seen in class, follows almost immediate from Lemma 1a.

To prove Lemma 1b we need another lemma.

Lemma 1d: Suppose h', h'' are supported on disjoint sets $T', T'' \subset [n]$ respectively, and A is $(|T'| + |T''|, \varepsilon_0)$-RIP. Then

$$|\langle Ah', Ah'' \rangle| \le \varepsilon_0 \|h'\|_2 \|h''\|_2.$$

Exer: Prove this lemma.

Hint: First assume WLOG that h', h'' are unit vectors. Then apply the formula $\|u+v\|_2^2 - \|u-v\|_2^2 = 4\langle u, v \rangle$ to $u = Ah'$ and $v = Ah''$.

Proof of Lemma 1b: Was seen in class. The idea is to analyze the norm of $Ah_{T_0 \cup T_1}$ (instead of that of $h_{T_0 \cup T_1}$) to show

$$(1 - \varepsilon)\|h_{T_0 \cup T_1}\|_2^2 \le \|Ah_{T_0 \cup T_1}\|_2^2 \le \varepsilon\sqrt{2}\|h_{T_0 \cup T_1}\|_2 \sum_{j \ge 2} \|h_{T_j}\|_2,$$

then plug in Lemma 1c, and rearrange.