Sublinear Time and Space Algorithms 2020B – Lecture 6
Compressed Sensing, RIP matrices and Basis Pursuit*

Robert Krauthgamer

1 Application 2: Compressed Sensing (cont'd)

Approximately sparse vector: We will now prove an even more general result.

For $z \in \mathbb{R}^n$, denote by $z_{top(k)}$ the vector z after zeroing all *but* the k heaviest entries (largest in absolute value), breaking ties arbitrarily. Notice this vector is the "best" k-sparse approximation to z. Similarly, denote by $z_{tail(k)} \in \mathbb{R}^n$ the vector z after zeroing the k heaviest entries. Then $z_{tail(k)} = z - z_{top(k)}$ is the "error" of approximating z by a k-sparse vector.

Theorem 1 [Cormode and Muthukrishnan, 2006]: CountMin++ with parameter $\alpha = \varepsilon/k$ can be used to recover a vector $x' \in \mathbb{R}^n$ that with high probability satisfies

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

In fact, $x' = \hat{x}_{top(k)}$ and is thus k-sparse. (Recall $\hat{x} \in \mathbb{R}^n$ is the estimate of algorithm CountMin.)

The above condition is usually called an ℓ_1/ℓ_1 guarantee.

Remark 1: Observe that if x is k-sparse, then this guarantees exact recovery. In general, it guarantees the output's "accuracy" (distance from true x) is comparable to the best k-sparse vector.

Remark 2: While in point queries we bounded the error in each coordinate separately, the above guarantee bounds the total error (over all coordinates).

Remark 3: Different constructions achieve/optimize for other guarantees like different norms, deterministic recovery, small explicit description of A, or fast recovery time. Often, the optimal number of measurements is $O(k \log(n/k))$ (ignoring dependence on ε).

Lemma 1a: CountMin++ with parameter α computes, with high probability, estimates $\hat{x}_i \in x_i \pm \alpha\|x_{tail(k)}\|_1$, i.e.,

$$\|x - \hat{x}\|_\infty \le \alpha\|x_{tail(k)}\|_1.$$

Exer: Prove this lemma.

*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.

Hint: Show that with high probability, both (a) coordinate i will not collide with the k (other) heaviest coordinates and (b) the contribution from the rest (tail) is comparable to the expectation.

Lemma 1b: If $\|x - \hat{x}\|_\infty \le \alpha \|x_{tail(k)}\|_1$ then $\|x - \hat{x}_{top(k)}\|_1 \le (1 + 3k\alpha)\|x_{tail(k)}\|_1$.

Proof of lemma: Let z_S denote the vector z after zeroing all coordinates outside $S \subset [n]$.

Let $\hat{T} \subset [n]$ be the indices of the k heaviest coordinates in $\hat{x}$, then by definition $\hat{x}_{\hat{T}} = \hat{x}_{top(k)} = x'$.

Let $T \subset [n]$ be the indices of the k heaviest coordinates in x, hence $x_T = x_{top(k)}$.

Denote the upper bound we have for every coordinate by $B := \alpha \|x_{tail(k)}\|_1$.

$$\begin{aligned}
\|x - x'\| &= \|x_{\hat{T}} - \hat{x}_{\hat{T}}\| + \|x_{\hat{T}^c} - 0\| && \text{separate coordinates of } \hat{T} \\
&= \|x_{\hat{T}} - \hat{x}_{\hat{T}}\| + \|x\| - \|x_{\hat{T}}\| \\
&\le |\hat{T}| \cdot B + \|x\| - \|\hat{x}_{\hat{T}}\| + |\hat{T}| \cdot B && \text{by } x \approx \hat{x} \text{ on } \hat{T} \\
&= 2|\hat{T}| \cdot B + \|x\| - \|\hat{x}_{\hat{T}}\| \\
&\le 2|\hat{T}| \cdot B + \|x\| - \|\hat{x}_T\| && \hat{T} \text{ is heaviest in } \hat{x} \\
&\le 2|\hat{T}| \cdot B + \|x\| - \|x_T\| + |T| \cdot B && \text{by } \hat{x} \approx x \text{ on } T \\
&\le (2|\hat{T}| \cdot \alpha + 1 + |T| \cdot \alpha)\|x_{tail(k)}\|.
\end{aligned}$$

QED.

Exer: Can you extend the above sparse recovery to ℓ_2/ℓ_2 guarantee by using CountSketch (instead of CountMin)? How many measurements would it require?

2 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

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

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 that the columns of A_S are linearly independent.

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 $i \neq j$).

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.

3 Compressed Sensing via Basis Pursuit

Theorem 2 [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

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

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 $x^* = \tilde{x}_{top(k)}$:

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

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

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

* 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 $h_{T_0^c}$, and similarly for $T_2, T_3, \ldots$. Overall, $T_0, T_1, T_2, \ldots$ is a partition of $[n]$ into groups size k each (except maybe the last one).

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

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

The proof will be completed by the following two lemmas.

QED

Lemma 2a: $\|h_{T_0 \cup T_1}\|_2 \leq O(1/\sqrt{k})\|x_{T_0^c}\|_1$.

Lemma 2b: $\|h_{(T_0 \cup T_1)^c}\|_2 \leq O(1/\sqrt{k})\|x_{T_0^c}\|_1 + \|h_{T_0 \cup T_1}\|_2$.

We start by proving (next week) a strengthening of Lemma 2b.

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