Sublinear Time and Space Algorithms 2020B – Lecture 7
Basis Pursuit (cont'd) and Iterative Hard Thresholding*

Robert Krauthgamer

1 Compressed Sensing via Basis Pursuit (cont'd)

Last time we started proving the theorem below, but it remained to prove the two main lemmas below.

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

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

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

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

Proof of Lemma 2b+: The first inequality follows from $h_{(T_0\cup T_1)^c} = \sum_{j\ge 2} h_{T_j}$ and the triangle inequality.

The second inequality was seen in class using the so-called "shelling argument", and then using that $\tilde{x} = x - h$ is a minimizer of the LP to expand $\|x\|_1 \ge \|\tilde{x}\|_1$.

To prove Lemma 2a we need another lemma.

Lemma 2d: 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 $v = Ah''$.

*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.

Proof of Lemma 2a: Was seen in class. The idea is to analyze the norm of $Ah_{T_0 \cup T_1}$ (instead of that of $h_{T_0 \cup T_1}$), using Lemma 2d, 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 2b+, and rearrange.

2 Iterative Hard Thresholding (IHT)

We will now see a different model of Compressed Sensing, where the error/noise is introduced after the measurement.

Theorem 3: Let $A \in \mathbb{R}^{m \times n}$ be $(3k, \varepsilon)$-RIP for $\varepsilon < 0.1$. Then given $y = Ax + e$ for an (unknown) k-sparse vector $x \in \mathbb{R}^n$ and some noise vector $e \in \mathbb{R}^n$, one can recover in polynomial time an estimate $\hat{x}$ such that $\|\hat{x} - x\|_2 \le O(1)\|e\|_2$.

Henceforth, all norms are ℓ_2 norms.

Basic intuition: The algorithm initially computes $z = A^T y$, and takes $z_{top(k)}$.

Why is this effective? We expect that $z = A^T Ax + A^T e \approx x$, because $A^T Ax \approx x$ and $A^T e$ should be small noise. We will give a formal bound in Lemma 3a below.

The error is then reduced via iterations on the "residual error" in x.

Algorithm IHT:

1. init: $z^{(0)} \leftarrow A^T y$, then let $x^{(0)} \leftarrow z^{(0)}_{top(k)}$
2. for $t = 1, \ldots, l = O(\log \frac{\|x\|}{\|e\|})$:
3. compute $z^{(t)} \leftarrow x^{(t-1)} + A^T(y - Ax^{(t-1)})$, then let $x^{(t)} \leftarrow z^{(t)}_{top(k)}$.
4. output $\hat{x} = x^{(t)}$

Lemma 3a (initialization):

$$\|x^{(0)} - x\| \le \tfrac{1}{4}\|x\| + 2\|e\|.$$

Lemma 3b (iterative improvement): For every $t \ge 1$,

$$\|x^{(t)} - x\| \le \tfrac{1}{4}\|x^{(t-1)} - x\| + 5\|e\|.$$

Proof of Theorem 3: As discussed in class, it follows easily from Lemmas 3a and 3b.

Lemma 3c: Let $S \supset \mathrm{supp}(x)$, $|S| = 3k$. Then

$$\|(z^{(0)} - x)_S\| \le \varepsilon\|x\| + 2\|e\|.$$

Proof: Since $Ax = A_S x_S$ and since A is $(3k,\varepsilon)$-RIP,

$$\begin{aligned}
\|(z-x)_S\| &= \|A_S^T(Ax+e) - x_S\| \\
&\le \|(A_S^TA_S - I)x_S\| + \|A_S^Te\| && \text{(triangle inequality)}\\
&\le \|A_S^TA_S - I\|\|x_S\| + \|A_S^T\|\|e\| && \text{(operator norm)}\\
&\le \varepsilon\|x\| + 2\|e\|, && \text{(RIP)}
\end{aligned}$$

where we bounded $\|A_S^T\| = \|A_S\| = \sup\{\|A_S v\| : \|v\| = 1\} \le (1+\varepsilon)^{1/2} \le 2$.

Lemma 3d: Let $z \in \mathbb{R}^n$ and let $T \subset [n]$ be its k heaviest coordinates. Then

$$\|z_T - x\|^2 \le 5\|(z-x)_{T\cup \mathrm{supp}(x)}\|^2.$$

Remark: It actually holds for every $z \in \mathbb{R}^n$, not only for $z^{(0)} = A^T y$.

Proof: Denote $H = \mathrm{supp}(x)$.

Coordinates $i \in T \cap H$ contribute $(z_i - x_i)^2$ to the LHS, and 5 times that to RHS.

Coordinates $i \notin T \cup H$ contribute 0 to LHS, and nonnegatively to RHS.

Now pair each $i \in H \setminus T$ with $j \in T \setminus H$ ordered by magnitude, then $|z_i| \le |z_j|$. By considering what each pair contributes to each side, it suffices to show $x_i^2 + z_j^2 \le 5[(z_i - x_i)^2 + z_j^2]$.

If $|z_i| > |x_i|/2$, then $x_i^2 \le 4z_i^2 \le 4z_j^2$ and we're done.

Otherwise $|z_i| \le |x_i|/2$, then $5(x_i - z_i)^2 \ge 5(x_i/2)^2$ and we're done.

QED

Proof of Lemma 3a: Recall $z^{(0)} = A^T y$, and let $T \subset [n]$ be its k heaviest coordinates. Then

$$\begin{aligned}
\|z_T^{(0)} - x\| &\le \sqrt{5}\ \|(z^{(0)} - x)_{T\cup\mathrm{supp}(x)}\| && \text{(Lemma 3d)}\\
&\le \sqrt{5}\ [\varepsilon\|x\| + 2\|e\|] && \text{(Lemma 3c)}\\
&\le \tfrac{1}{4}\|x\| + 5\|e\|.
\end{aligned}$$

QED

Proof of Lemma 3b: We did not have time in class, but here it is.

For sake of analysis, consider a "hypothetical" input where we subtract the previous iteration:

$$\begin{aligned}
&x' = x - x^{(t-1)} \quad \Rightarrow \quad \mathrm{supp}(x') \subseteq \mathrm{supp}(x) \cup \mathrm{supp}(x^{(t-1)}) \quad \text{(has size } \le 2k)\\
&y' = Ax' + e \quad \Rightarrow \quad y' = A(x - x^{(t-1)}) + e = y - Ax^{(t-1)} \quad \text{(line 3 uses this } y')\\
&z' = A^T y'.
\end{aligned}$$

Using this notation, we can rewrite line 3 as $z^{(t)} \leftarrow x^{(t-1)} + z'$, and

$$z^{(t)} - x = x^{(t-1)} + z' - x = z' - x'.$$

Analogously to the proof of Lemma 3a:

$$\begin{aligned}
\|x^{(t)} - x\| &= \|z^{(t)}_{T^{(t)}} - x\| && \text{(denote } T^{(t)} = \mathrm{supp}(x^{(t)}) \text{)} \\
&\leq \sqrt{5}\ \|(z^{(t)} - x)_{T^{(t)} \cup \mathrm{supp}(x)}\| && \text{(Lemma 3d for } z^{(t)} \text{)} \\
&\leq \sqrt{5}\ \|(z' - x')_{T^{(t)} \cup \mathrm{supp}(x) \cup \mathrm{supp}(x^{(t-1)})}\| && \text{(rewrite as above)} \\
&\leq \sqrt{5}\ [\varepsilon \|x'\| + 2\|e\|] && \text{(Lemma 3c for } x', z' \text{)} \\
&\leq \tfrac{1}{4}\|x - z_T\| + 5\|e\|.
\end{aligned}$$

QED

Theorem 4 [L_1-minimization Algorithm]: A guarantee similar to Theorem 3 (using RIP matrix) can be obtained by setting $b \geq \|e\|$ and solving the convex program

$$\hat{x} = \min\{\|z\|_1 : \|Az\|_2 \leq b\}.$$

We will not see the proof.