Special Topics in Theoretical Computer Science February 14, 2019

Lecture 11: Compressed Sensing using ℓ_1 minimization

Huy L. Nguyen *Scribe: Akshar Varma*

Overview

In the last lecture we started looking at Compressed Sensing, in this lecture we show that solving an ℓ_1 norm minimization problem is sufficient for sparse signal recovery. We prove that we can recover any sparse signal using only a few linear measurements and that this can be done in polynomial time. This method of sparse recovery is sometimes called the Basis Pursuit method.

1 Compressed Sensing

The original signal we are interested in recovering is $x \in \mathbb{R}^n$ which is k-sparse (only has k non-zeroes) and may have noise added to it. Our goal is to find an $\hat{x}$ such that:

$$\|\hat{x} - x\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \cdot \|x - x^k\|_1$$

where x^k consists of only the top k coordinates of x in absolute value. Note that the ℓ_1 norm term on the RHS of the inequality denotes the optimal error possible. This means that if there is no noise then we wish to perfectly recover the k-sparse signal.

Definition 1 (Restricted Isometry Property (RIP)) $\Pi \in \mathbb{R}^{m\times n}$ *satisfies* (ϵ, k)*-RIP if* $\forall x$ *that are* k*-sparse,*

$$(1-\epsilon)\|x\|_2^2 \leq \|\Pi x\|_2^2 \leq (1+\epsilon)\|x\|_2^2$$

We proved last time, using ϵ-nets and the JL lemma applied to $\binom{n}{k}$ k-dimensional subspaces of $\mathbb{R}^n$, that there exists such an RIP matrix Π for $m = O(\frac{k}{\epsilon^2}\log\frac{n}{k})$.

We now state Theorem 2, which is the main theorem of compressed sensing.

Theorem 2 *There is a poly-time algorithm which, given* Πx *for some* Π *satisfying the* $(\sqrt{2}-1, 2k)$*-RIP, can recover* $\hat{x}$ *such that*

$$\|\hat{x} - x\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right)\|x - x^k\|_1$$

Ideally we would like to solve the following optimization problem:

$$\min_{\Pi z = \Pi x} \|z\|_0$$

However, since this is an NP-hard non-convex optimization problem, we use the tightest convex relaxation of the problem:

$$\min_{\Pi z = \Pi x} \|z\|_1$$

This convex relaxation can actually be written as a linear program by rewriting z as a positive and a negative component, $z = z^+ - z^-$.

$$\begin{aligned} \min \quad & (z^+ + z^-) \\ \text{s.t.} \quad & \Pi(z^+ - z^-) = \Pi x \\ & z^+, z^- \geq 0 \end{aligned} \tag{1}$$

This allows us to use any of the polynomial time algorithms for linear programming to solve this optimization problem and get an optimal z^*. We will prove, in Theorem 3, that such a z^* satisfies the required bound for sparse signal recovery. Since finding the optimal solution z^* can be done in polynomial time, Theorem 2 becomes a straighforward corollary of Theorem 3.

Theorem 3 *If Π is $(\epsilon, 2k)$-RIP with $\epsilon \leq \sqrt{2} - 1$, then the optimal solution $z = x + h$ for the optimization problem 1 satisfies:*

$$\|h\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \|x - x^k\|_1$$

Proof. The core idea in the proof is to sort the coordinates of x and h by their absolute values and put them into groups of size k. We use the following notation:

- x_S is x with all coordinates outside S zeroed out.
- $\overline{S}$ is the complement of S.

Using that notation, we define a partition of the coordinates into sets of size at most k:

- T_0 is the set of indices of the k-largest coordinates of x in absolute value.
- T_1 is the set of indices of the k-largest coordinates of h in absolute value, after excluding the indices in T_0.
- For $i \geq 2$, T_i is the set of indices of the k-largest coordinates of h in absolute value, after excluding the indices $\bigcup_{j=0}^{j<i} T_j$.

The above partition of indices into sets of size k is sometimes called the shelling trick. A property of such a partition that we will rely on is the following: for all $j \in T_i$ and for any $k \in T_{i-1}$, it holds that $|x_j| \leq |x_k|$.

We bound $\|h\|_2$ by using the triangle inequality to split it into two terms and then show that both those terms are small:

$$\|h\|_2 \leq \|h_{T_0 \cup T_1}\|_2 + \|h_{\overline{T_0 \cup T_1}}\|_2$$

We will show that both the terms on the RHS are small by proving the following claims

Claim 1

$$\|h_{T_0\cup T_1}\|_2 \le O\left(\frac{1}{\sqrt{k}}\right)\|x_{\overline{T_0}}\|_1$$

Claim 2

$$\|h_{\overline{T_0\cup T_1}}\|_2 \le \|h_{T_0\cup T_1}\|_2 + O\left(\frac{1}{\sqrt{k}}\right)\|x_{\overline{T_0}}\|_1$$

We first prove Claim 2 using the following Lemma

Lemma 1

$$\sum_{j\ge 2}\|h_{T_j}\|_2 \le \frac{2}{\sqrt{k}}\|x_{\overline{T_0}}\|_1 + \|h_{T_0\cup T_1}\|_2$$

Proof. We rely on the shelling trick property mentioned earlier for proving this lemma

$$\begin{aligned}\sum_{j\ge 2}\|h_{T_j}\|_2 &\le \sum_{j\ge 2}\|h_{T_j}\|_\infty\sqrt{k}\\ &\le \sum_{j\ge 1}\|h_{T_j}\|_1\frac{\sqrt{k}}{k}\\ &= \frac{\|h_{\overline{T_0}}\|_1}{\sqrt{k}} \qquad (2)\end{aligned}$$

Now we use the optimality of $z = x + h$: for any vector y, $\|y\|_1 \ge \|z\|_1$ and in particular this is true for $y = x$.

$$\begin{aligned}\|x\|_1 &\ge \|z\|_1\\ &\ge \|(x+h)_{T_0}\|_1 + \|(x+h)_{\overline{T_0}}\|_1\end{aligned}$$

Using the triangle inequality

$$\|x_{T_0}\|_1 + \|x_{\overline{T_0}}\|_1 \ge \|x_{T_0}\|_1 - \|h_{T_0}\|_1 + \|h_{\overline{T_0}}\|_1 - \|x_{\overline{T_0}}\|_1$$

And then shuffling the terms

$$\begin{aligned}\|h_{\overline{T_0}}\|_1 &\le 2\|x_{\overline{T_0}}\|_1 + \|h_{T_0}\|_1\\ &\le 2\|x_{\overline{T_0}}\|_1 + \sqrt{k}\|h_{T_0}\|_2\\ &\le 2\|x_{\overline{T_0}}\|_1 + \sqrt{k}\|h_{T_0\cup T_1}\|_2 \qquad (3)\end{aligned}$$

Plugging in the value in Equation 3 into Equation 2 proves the lemma. □

Claim 2 can then be proved using the triangle inequality and Lemma 1.

$$\|h_{\overline{T_0\cup T_1}}\|_2 \le \sum_{j\ge 2}\|h_{T_j}\|_2 \le \frac{2}{\sqrt{k}}\|x_{\overline{T_0}}\|_1 + \|h_{T_0\cup T_1}\|_2$$

Lemma 2 *If vectors u and v are supported on disjoint sets S, T of coordinates respectively, and $|S|, |T| \le k$ then $|\langle \Pi u, \Pi v\rangle| \le \epsilon \|u\|_2 \|v\|_2$ if Π is $(\epsilon, 2k)$-RIP.*

Proof. Without loss of generality we can assume that u and v are unit vectors. We use the following identity

$$ab = \frac{1}{4}\left((a+b)^2 - (a-b)^2\right)$$

When applied on Πu and Πv and then using the RIP property we get:

$$\begin{aligned}\langle \Pi u, \Pi v\rangle &= \frac{1}{4}\Big(\|\Pi(u+v)\|_2^2 - \|\Pi(u-v)\|_2^2\Big) \\ &= \frac{1}{4}\Big(2(1+\epsilon) - 2(1-\epsilon)\Big) \\ &= \epsilon\end{aligned}$$

□

Using Lemma 2 we can prove Claim 1. Since $z = x + h$ satisfies the constraint $\Pi z = \Pi x$,

$$\Pi(x+h) = \Pi x \implies \Pi h = 0 \implies \Pi\left(h_{T_0 \cup T_1}\right) = -\Pi\left(h_{\overline{T_0 \cup T_1}}\right)$$

Since we want to bound $\|h_{T_0 \cup T_1}\|_2$, we square it, write it as an inner product and simplify using Lemma 2.

$$\begin{aligned}\|\Pi h_{T_0\cup T_1}\|_2^2 &= \langle \Pi h_{T_0\cup T_1}, \Pi h_{T_0\cup T_1}\rangle \\ &= \langle \Pi h_{T_0\cup T_1}, -\Pi h_{\overline{T_0\cup T_1}}\rangle \\ &= \sum_{j\ge 2}\left|\langle \Pi h_{T_0}, \Pi h_{T_j}\rangle\right| + \sum_{j\ge 2}\left|\langle \Pi h_{T_1}, \Pi h_{T_j}\rangle\right| \\ &\le \sum_{j\ge 2} \epsilon\cdot\left(\|h_{T_0}\|_2 + \|h_{T_1}\|_2\right)\cdot \|h_{T_j}\|_2\end{aligned}$$

Now we simplify using the fact that h_{T_0} and h_{T_1} are orthogonal. Then we use the RIP property and Lemma 1.

$$\begin{aligned}\|\Pi h_{T_0\cup T_1}\|_2^2 &\le \sum_{j\ge 2}\epsilon\sqrt{2}\|h_{T_0\cup T_1}\|_2\cdot\|h_{T_j}\|_2 \\ (1+\epsilon)\|h_{T_0\cup T_1}\|_2^2 &\le \sum_{j\ge 2}\epsilon\sqrt{2}\|h_{T_0\cup T_1}\|_2\cdot\|h_{T_j}\|_2 \\ (1+\epsilon)\|h_{T_0\cup T_1}\|_2^2 &\le \epsilon\sqrt{2}\|h_{T_0\cup T_1}\|_2\left(\|h_{T_0\cup T_1}\|_2 + O\left(\frac{1}{\sqrt{k}}\right)\|x_{\overline{T_0}}\|_1\right) \\ \|h_{T_0\cup T_1}\|_2 &\le \frac{\epsilon\sqrt{2}}{1-\epsilon}\left(\|h_{T_0\cup T_1}\|_2 + O\left(\frac{1}{\sqrt{k}}\right)\|x_{\overline{T_0}}\|_1\right) \\ \|h_{T_0\cup T_1}\|_2 &\le \frac{\frac{\epsilon\sqrt{2}}{1-\epsilon}}{1-\frac{\epsilon\sqrt{2}}{1-\epsilon}}\cdot O\left(\frac{1}{\sqrt{k}}\right)\|x_{\overline{T_0}}\|_1\end{aligned}$$

This proves Claim 1 and hence both the theorems as well. □

2 Note about RIP matrices

Last time we saw that the distributional JL lemma allowed us to prove the existence of a matrix with the RIP property. The reverse direction is also possible. That is, if there is a matrix M with the RIP property then we can construct a distribution on matrices satisfying the distrubutional JL lemma with $O(\epsilon)$ error and a failure probability of $2^{-\Omega(k)}$.

Note that we lose a little in this sequence of transformations. If we start with a distributional JL lemma with failure probability $2^{-\Omega(k\log(\frac{n}{k}))}$, get a RIP matrix and use that to come up with a distributional JL, then we lose the $\log(\frac{n}{k})$ factor and have a slightly worse failure probability than we started with.

3 More on Compressed Sensing

- We saw how to use LP-solvers for sprase recovery. But these methods can be slow since the LP has a lot of constraints. This can be improved by using other methods.
- Sometimes, the signal we are interested in may not be sparse in the basis in which we measure but in a different basis. In some cases, we can also recover the sparse signal in a different basis.
- We can also extend to cases where there is an error during or post measurement. This will cause our recovery algorithm to incur an additive error proportional to the post measurement error.
- More structure on the signal can help reduce the number of required measurements. For example, in an image, if we look at difference between neighbouring pixels, then that signal is sparse but it also has the additional property that all the non-zero entries are connected and form object boundaries in the image. Such information can help us improve the number of measurements needed for sparse recovery.