CS 229r: Algorithms for Big Data Fall 2015

Lecture 19 — Nov 5

Prof. Jelani Nelson *Scribe: Abdul Wasay*

1 Overview

In the last lecture, we started discussing the problem of compressed sensing where we are given a sparse signal and we want to reconstruct is using as little measurements as possible. We looked into how sparsity manifests itself in images which are approximately sparse in Haar wavelet basis.

In this lecture we will look at

- Recap a bit about compressive sensing.
- RIP and connection to incoherence
- Basis pursuit
- Krahmer-Ward theorem

2 Review from Last time

2.1 Compressed Sensing

In compressed sensing, we are given a "compressible" signal $x \in \mathbb{R}^n$, and our goal is use few linear measurements of x to approximately recover x. Here, a linear measurement of x is its dot product with another vector in $\mathbb{R}^n$. We can arrange m such linear measurements to form the rows of a matrix $\Pi \in \mathbb{R}^{m\times n}$, so the goal now becomes to approximately recover x from Πx using $m \ll n$.

Note that if $m < n$, then any Π has a non-trivial kernel, so we have no hope of exactly recovering every $x \in \mathbb{R}^n$. This motivates our relaxed objective of only recovering *compressible* signals.

So what exactly do we mean by compressible? A compressible signal is one which is (approximately) sparse in some basis – but not necessarily the standard basis. Here an approximately sparse signal is a sum of a sparse vector with a low-weight vector.

2.2 Algorithmic Goals

The compressed sensing algorithms we discuss will achieve the following. If x is actually sparse, we will recover x exactly in polynomial time. And if x is only approximately sparse, then we will recover x approximately, again in poly-time.

More formally, we seek to meet "ℓ_p/ℓ_q guarantees": Given Πx, we will recover $\tilde{x}$ such that

$$\|x - \tilde{x}\|_p \leq C_{k,p,q} \min_{\|y\|_0 \leq k} \|y - x\|q,$$

where the ℓ_0-norm of a vector is the count of its non-zero coordinates. Observe that the minimizer y in the statement above picks out the largest (in absolute value) k coordinates of x and zeroes out the rest of them. Also, note that the right-hand side is zero when x is actually k-sparse.

3 Main Result

Theorem 1. *There is a polynomial-time algorithm which, given Πx for $\Pi \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, can recover $\tilde{x}$ such that*

$$\|x - \tilde{x}\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1$$

where $x_{tail(k)}$ is x with its top k coordinates zeroed out.

3.1 Exact recovery in the sparse case

As a first step toward proving the theorem, let's examine what we need to recover x exactly when we actually have $\|x\|_0 \leq k$. Information- theoretically, it's necessary and sufficient to have $\Pi x \neq \Pi x'$ whenever $x \neq x'$ are both k-sparse. This is equivalent to requiring any $2k$-sparse vector to lie outside $\ker \Pi$, i.e., requiring each restriction Π_S of Π to the columns in a set S to have full column rank for every $S \subseteq [n]$ with $|S| \leq 2k$.

How can we use this characterization to recover x given $y = \Pi x$ when Π has this property? One way is to find the minimizer z in

$$\begin{aligned} \min_{z\in\mathbb{R}^n} \quad & \|z\|_0 \\ \text{s.t.} \quad & \Pi z = y. \end{aligned}$$

Unfortunately, this optimization problem is NP-hard to solve in general [GJ79, Problem MP5]. In what follows, we will show that with an additional constraint on Π, we can approximately solve this optimization problem using linear programming.

4 RIP Matrices

Definition 2. *A matrix $\Pi \in \mathbb{R}^{m\times n}$ satisfies the (ε, k)- restricted isometry property (RIP) if for all k-sparse vectors x:*

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

Equivalently, whenever $|S| \leq k$, we have

$$\|\Pi_S^T \Pi_S - I_k\|_2 \leq \varepsilon.$$

5 Calculating RIP matrices

RIP matrices are obtainable from the following methods:

- Use **JL** to preserve each of the k-dimensional subspace. This can be done by applying JL to the requisite $\binom{n}{k}e^{O(k)}$ vectors yields, by Stirling's approximation,

$$m \lesssim \frac{1}{\varepsilon^2} k \log\left(\frac{n}{k}\right).$$

- Use **incoherent matrices**. The good thing about incoherent matrices is that they are explicit from codes as we have looked in problem set 1.
- From **first principles**. A matrix Π might not satisfy JL, but might still preserve the norms of k-sparse vectors. For example, we can take Π to sample m rows from a Fourier matrix. Recall that for the FJLT, we needed to subsequently multiply by a diagonal sign matrix, but there is no need to do so in the particular case of sparse vectors.

We will focus on the second method.

Recall 3. *A matrix Π is $\alpha-$coherent if:*

- $\|\Pi^i\|_2 = 1$ *for all i, and*
- $|\langle \Pi^i, \Pi^j \rangle| \leq \alpha$ *for all $i \neq j$.*

Claim 4. *Incoherent matrices can be used to explicitly construct RIP.*

We will use the **Gershgorin Circle Theorem** to prove the above claim.

Lemma 5. *Given a matrix A, all its eigenvalues, lie within a complex disk of radius $\sum_{j\neq i} |a_{ij}|$.*

Proof. Let x be an eigenvector of A with corresponding eigenvalue λ. Let i be an index such that $|x_i| = \|x\|_\infty$. Then $(Ax)_i = \lambda x_i$ so

$$\begin{aligned}\lambda x_i = \sum_{j=1}^{n} a_{ij}x_j \Rightarrow |(\lambda - a_{ii})x_i| &= \left|\sum_{j\neq i} a_{ij}x_j\right| \\ \Rightarrow |\lambda - a_{ii}| &\leq \sum_{j\neq i}\left|\frac{a_{ij}x_j}{x_i}\right| \leq \sum_{j\neq i}|a_{ij}|\end{aligned}$$

by our choice of i. □

Theorem 6. *If Π is $(\frac{\varepsilon}{k})$ incoherent, it implies that it is (ε, k)-RIP.*

Proof. Suppose we have an $\alpha-$incoherent matrix where $\alpha = \frac{\varepsilon}{k}$. Let us analyze A=$(\Pi I_s)^T(\Pi I_s)$ for some $|s| \leq k$. Notice that A is a symmetric matrix and it has real eigenvalues. Also, $a_{ii} = 1$ and $a_{ij} = \alpha = \frac{\varepsilon}{k}$. Therefore, the eigenvalues of A lie in the interval of radius:

$$\sum_{j \neq i} |\langle \Pi_S^i, \Pi_S^j \rangle| \leq \alpha(k-1)$$

We set $\alpha = \frac{\varepsilon}{k}$ and get an (ε, k)-RIP matrix. $\square$

6 Basis Pursuit OR RIP to Recovery

Theorem 7. *If Π is $(\varepsilon_{2k}, 2k)$-RIP with $\varepsilon_{2k} \leq \sqrt{2} - 1$, and $\tilde{x} = x + h$ is the solution to the "basis pursuit" linear program*

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & \|z\|_1 \\ \text{s.t.} \quad & \Pi z = \Pi x, \end{aligned}$$

then

$$\|h\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1.$$

Remark: A *linear program* (LP) is an optimization problem in which one seeks to optimize a linear objective function subject to linear constraints. The above problem is indeed a linear program with polynomially many variables and constraints, since it is equivalent to

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \quad & \sum_i y_i \\ \text{s.t.} \quad & \Pi z = \Pi x, \\ & z_i \leq y_i \quad \forall i, \\ & -z_i \leq y_i \quad \forall i. \end{aligned}$$

It is known (e.g. via Khachiyan's analysis of the ellipsoid method) that LPs can be solved in polynomial time.

We will now present a proof along the lines of [Candes08].

Proof. First, we define some notation.

For a vector $x \in \mathbb{R}^n$ and a set $S \subseteq [n]$, let x_S be the vector with all of its coordinates outside of S zeroed out. We will use T_i^c to indicate the complement of T_i

- Let $T_0 \subseteq [n]$ be the indices of the largest (i absolute value) k coordinates of x.
- Let T_1 be the indices of the largest k coordinates of $h_{T_0^c} = h_{tail(k)}$.
- Let T_2 be the indices of the *second* largest k coordinates of $h_{T_0^c}$.
- ...and so forth, for $T_3, \ldots$

By the triangle inequality, we can write

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

Our strategy for bounding h will be to show:

1. $$\|h_{(T_0 \cup T_1)^c}\|_2 \leq \|h_{T_0 \cup T_1}\|_2 + O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1.$$

2. $$\|h_{T_0 \cup T_1}\|_2 \leq O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1)$$

Both parts rely on the following lemma.

Lemma 8.

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

Proof. We first get an upper bound on the left-hand side by applying a technique known as the "shelling trick."

$$\begin{aligned} \sum_{j \geq 2} \|h_{T_j}\|_2 &\leq \sqrt{k} \sum_{j \geq 2} \|h_{T_j}\|_\infty \\ &\leq \frac{1}{\sqrt{k}} \sum_{j \geq 2} \|h_{T_{j-1}}\|_1 \\ &\leq \frac{1}{\sqrt{k}} \|h_{T_0^c}\|_1. \end{aligned} \tag{1}$$

The first inequality holds because each h_{T_j} is k-sparse, and the second holds because the size of every term in h_{T_j} is bounded from above by the size of every term in $h_{T_{j-1}}$.

Now since $\tilde{x} = x + h$ is the minimizer of the LP, we must have

$$\begin{aligned} \|x\|_1 &\geq \|x + h\|_1 \\ &= \|(x+h)_{T_0}\|_1 + \|(x+h)_{T_0^c}\|_1 \\ &\geq \|x_{T_0}\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 - \|x_{T_0^c}\|_1 \end{aligned}$$

by two applications of the reverse triangle inequality. Rearranging, we obtain

$$\begin{aligned} \|h_{T_0^c}\|_1 &\leq \|x\|_1 - \|x_{T_0}\|_1 + \|h_{T_0}\|_1 + \|x_{T_0^c}\|_1 \\ &= 2\|x_{T_0^c}\|_1 + \|h_{T_0}\|_1 \\ &\leq 2\|x_{T_0^c}\|_1 + \sqrt{k}\|h_{T_0}\|_2 && \text{by Cauchy- Schwarz} \\ &\leq 2\|x_{T_0^c}\|_1 + \sqrt{k}\|h_{T_0 \cup T_1}\|_2 \end{aligned}$$

Combining this upper bound with Inequality (1) yields the claim.

□

Returning to the main proof, let us first upper bound the size of $h_{(T_0 \cup T_1)^c}$. We get:

$$\begin{aligned}
\|h_{(T_0\cup T_1)^c}\|_2 &= \left\| \sum_{j\geq 2} h_{T_j} \right\|_2 \\
&\leq \sum_{j\geq 2} \|h_{T_j}\|_2 \\
&\leq \|h_{T_0\cup T_1}\|_2 + \frac{2}{\sqrt{k}}\|x_{T_0^c}\|_1 && \text{by the claim} \\
&= \|h_{T_0\cup T_1}\|_2 + \frac{2}{\sqrt{k}}\|x_{tail(k)}\|_1 .
\end{aligned}$$

Now to bound the size of $h_{T_0\cup T_1}$, we need another lemma.

Lemma 9. *If x, x' are supported on disjoint sets T, T' respectively, where $|T| = k$ and $|T'| = k'$, then*

$$|\langle \Pi x, \Pi x' \rangle| \leq \varepsilon_{k+k'} \|x\|_2 \|x'\|_2,$$

where Π is $(\varepsilon_{k+k'}, k+k')$-RIP.

Proof. We can assume WLOG that x, x' are unit vectors. Write

$$\|\Pi x + \Pi x'\|_2^2 = \|\Pi x\|_2^2 + \|\Pi x'\|_2^2 + 2\langle \Pi x, \Pi x'\rangle, \quad \text{and}$$
$$\|\Pi x - \Pi x'\|_2^2 = \|\Pi x\|_2^2 + \|\Pi x'\|_2^2 - 2\langle \Pi x, \Pi x'\rangle.$$

Taking the difference gives

$$\begin{aligned}
|\langle \Pi x, \Pi x'\rangle| &= \frac{1}{4}\left| \|\Pi(x+x')\|_2^2 - \|\Pi(x-x')\|_2^2 \right| \\
&\leq \frac{1}{4}((1+\varepsilon_{k+k'})\|x+x'\|_2^2 - (1-\varepsilon_{k+k'})\|x-x'\|_2^2) \\
&= \frac{1}{4}((1+\varepsilon_{k+k'})2 - (1-\varepsilon_{k+k'})2) \\
&= \varepsilon_{k+k'}
\end{aligned}$$

since $x \pm x'$ are $(k+k')$-sparse, and x, x' are disjointly supported. This proves the lemma. □

To bound the size of $h_{T_0\cup T_1}$, first observe that

$$\Pi h_{T_0\cup T_1} = \Pi h - \sum_{j\geq 2} \Pi h_{T_j} = -\sum_{j\geq 2} \Pi h_{T_j}$$

since $h \in \ker \Pi$. Therefore,

$$\|\Pi h_{T_0 \cup T_1}\|_2^2 = -\sum_{j \geq 2} \langle \Pi h_{T_0 \cup T_1}, \Pi h_{T_j} \rangle \leq \sum_{j \geq 2} (|\langle \Pi h_{T_0}, \Pi h_{T_j} \rangle| + |\langle \Pi h_{T_1}, \Pi h_{T_j} \rangle|).$$

By Lemma 9, each summand is at most

$$\varepsilon_{2k}(\|h_{T_0}\|_2 + \|h_{T_1}\|_2)\|h_{T_j}\|_2 \leq \varepsilon_{2k}\sqrt{2}\|h_{T_0 \cup T_1}\|_2\|h_{T_j}\|_2.$$

Thus

$$\begin{aligned}
(1 - \varepsilon_{2k})\|h_{T_0 \cup T_1}\|_2^2 &\leq \|\Pi h_{T_0 \cup T_1}\|_2^2 \\
&\leq \varepsilon_{2k}\sqrt{2}\|h_{T_0 \cup T_1}\|_2 \sum_{j \geq 2} \|h_{T_j}\|_2 \\
&\leq \varepsilon_{2k}\sqrt{2}\|h_{T_0 \cup T_1}\|_2 \left(\frac{2}{\sqrt{k}} \|x_{T_0^c}\|_1 + \|h_{T_0 \cup T_1}\|_2 \right)
\end{aligned}$$

by Claim 8. Cancelling a factor of $\|h_{T_0 \cup T_1}\|_2$ from both sides and rearranging gives

$$\|h_{T_0 \cup T_1}\|_2 \leq \frac{\varepsilon_{2k} 2\sqrt{2}}{(1 - \varepsilon_{2k} - \varepsilon_{2k}\sqrt{2})\sqrt{k}} \|x_{T_0^c}\|_1 = O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1.$$

Putting everything together:

$$\begin{aligned}
\|h\|_2 &\leq \|h_{(T_0 \cup T_1)^c}\|_2 + \|h_{T_0 \cup T_1}\|_2 \\
&\leq 2\|h_{T_0 \cup T_1}\|_2 + \frac{2}{\sqrt{k}} \|x_{tail(k)}\|_1 \\
&\leq O\left(\frac{1}{\sqrt{k}}\right) \|x_{tail(k)}\|_1.
\end{aligned}$$

□

7 Krahmer and Ward

Theorem 10. *Let* $\Pi \in \mathbb{R}^{m \times n}$ *be a matrix satisfying the* $(\varepsilon, 2k)$*-RIP. Let* $\sigma \in \{+1, -1\}^n$ *be uniformly random and* D_σ *the diagonal matrix with* σ *on the diagonal. Then* ΠD_σ *satisfies the* $(O(\varepsilon), 2^{-\Omega(k)})$*-distributional JL property.*

This theorem states that given a matrix satisfying the RIP property, we can construct a distribution on matrices satisfying the JL property.

References

[BDFKK11] Jean Bourgain, Stephen Dilworth, Kevin Ford, Sergei Konyagin, and Denka Kutzarova. Breaking the k^2 Barrier for Explicit RIP Matrices. In *STOC*, pages 637–644, 2011.

[Candes08] Emmanuel Candès. The restricted isometry property and its implications for compressed sensing. *C.* R. Acad. Sci. Paris, Ser. I 346:589–592, 2008.

[CRT04] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[Don04] David Donoho. Compressed Sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[GJ79] Michael R. Garey, David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: Freeman, 1979.