15-750: Graduate Algorithms Mar 27, 2017

Lecture 24: Compressed Sensing and JL property

Lecturer: David Witmer *Scribe: Ziye Tang*

1 Recap

Recall the JL Lemma from last class. We have t points in $\mathbb{R}^n : \{x_1, \ldots, x_t\}$. We want to map all points to R^m for $m \ll n$ such that all distances are 'approximately' maintained. The idea is made clear in the lemma:

Lemma 1.1. *Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $X = \{x_1, \ldots, x_t\}$, then there exists a map $A : \mathbb{R}^n \to \mathbb{R}^m$ with $m = O(\frac{\log n}{\epsilon^2})$ such that $(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|A(x_i) - A(x_j)\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2$. Furthermore, we can take A to be linear.*

Here is one way to construct A: let M be a $m \times n$ matrix such that each entry M_{ij} is an independent subgaussian random variable with mean 0 and variance 1. Let $A = \frac{1}{\sqrt{m}}M$. We showed the following lemma with respect to such A:

Lemma 1.2. *Let $\epsilon \in (0, \frac{1}{2})$. For every unit vector $x \in \mathbb{R}^n$ and some constant $c > 0$, $\mathbb{P}[1 - \epsilon \leq \|Ax\|_2^2 \leq 1 + \epsilon] \geq 1 - e^{-c\epsilon^2 m}$*

Now we generalize and say that a distribution D has JL property if D is a distribution over $\mathbb{R}^{m\times n}$ and lemma 1.2 holds for $A \in \mathbb{R}^{m\times n}$ drawn from distribution D. We will explore the connection between JL property and compressed sensing in this lecture. The lecture is based on Matousek's notes.

2 Compressed Sensing

Definition 2.1. For vector x, let x_i be its ith entry. Define $supp(x) := \{i : x_i \neq 0\}$.

Applications of compressed sensing can be found in section 2 here. The main question we ask is the following: given $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, find x such that $Ax = b$ and $|supp(x)| \leq r$. The following theorem gives a necesssary and sufficient condition for the existence of such a solution:

Theorem 2.2. *$Ax = b$ has at most one solution x with $|supp(x)| \leq r$ if and only if every set of $2r$ columns of A is linearly independent.*

Proof. The proof is based on definitions of linear independence.

'if': suppose by contradiction that $Ax = b$ has two sparse solutions x_1, x_2. Let $\Delta x = x_1 - x_2$. Observe $supp(\Delta x) \leq 2r, A\Delta x = 0$ which is a contradiction.

'only if': suppose there are $2r$ columns that are linealy dependent, this means that $\exists y$, such that $|supp(y)| \leq 2r, Ay = 0$. Write $y = x_1 - x_2$ where $|supp(x_i)| \leq r, \forall i = 1, 2$, which is a contradiction. □

Despite the nice characterization, it is shown that finding such a sparse solution is NP-hard. The following idea called 'basis pursuit' is a heuristic to find a sparse solution. We consider the following optimization problem $\min\{\|x\|_1 : Ax = b\}$. Moreover it is easy to see that it is equivalent to the following LP.

$$\begin{aligned} \min \quad & \sum_i u_i \\ \text{s.t.} \quad & Ax = b \\ & -u_i \le x_i \le u_i \\ & u \ge 0 \end{aligned}$$

The above heuristic works well in practice. Can we prove that it works well for some natural class of matrices A?

Definition 2.3. Matrix $A \in \mathbb{R}^{m \times n}$ is BP-exact for sparsity r if for all $b \in \mathbb{R}^m$ such that $Ax = b$ has a unique sparse solution $\widetilde{x}$ with $|supp(\widetilde{x})| \le r$, the BP LP has a unique optimum $\widetilde{x}$.

Here is the main result we shall prove:

Theorem 2.4 (Donoho; Candes-Tao; Rudelson-Vershynin)**.** *There exists constants $C, c_1 > 0$ such that if n, m, r are integers with $1 \le r \le n/C$ and $m \ge Cr \log \frac{n}{r}$ and if $A \in \mathbb{R}^{m \times n}$ is a random matrix drawn from a distribution D satisfying JL property, then the following holds:*

$$\mathbb{P}[A \textit{ is BP-exact for sparsity } r] \ge 1 - e^{-c_1 m}$$

3 Proof

We outline the proof following Matousek's lecture notes:

1. Define restricted isometry property (RIP)
2. Show RIP $\implies$ BP-exactness
3. Show JL property $\implies$ RIP

3.1 Restricted isometry property

Definition 3.1. Let $A \in \mathbb{R}^{m \times n}$ and $\epsilon \in (0, 1)$. A is an ϵ-almost isometry if for all unit vector $x \in \mathbb{R}^n$, $1 - \epsilon \le \|Ax\|_2 \le 1 + \epsilon$

Definition 3.2. A is a t-restricted ϵ-almost isometry if for all unit vector $x \in \mathbb{R}^n$ such that $|supp(x)| \le t$, $1 - \epsilon \le \|Ax\|_2 \le 1 + \epsilon$.

Lemma 3.3 (RIP $\implies$ BP-exactness)**.** *There exists a constant $\epsilon_0 > 0$ such that if A is $3r$-restricted ϵ_0-almost isometry, then A is BP-exact for sparsity r.*

Proof. Suppose there exists $\Delta \ne 0$ such that $\widetilde{x} + \Delta$ is a solution to $Ax = b$ and $\|\widetilde{x} + \Delta\|_1 \le \|\widetilde{x}\|_1$. Since $A\widetilde{x} = b$, we know $A\Delta = 0$. Note if A is almost isometry, then we would have a contradiction from $A\Delta = 0, \Delta \ne 0$. Since we only have control of small subsets of columns in A, we need to estimate more carefully. Below we shall show, roughly speaking, that $supp(\widetilde{x})$ accounts for most L_1 and L_2 norm. For notation ease, let $S := supp(\widetilde{x})$ and Δ_S be the components of Δ indexed by S.

Below is the claim w.r.t. 1-norm, where we use that $\|\widetilde{x}\|_1 \ge \|\widetilde{x} + \Delta\|_1$

Claim 3.4. $\|\Delta_S\| \geq \|\Delta_{\overline{S}}\|$, *where* $\overline{S} = [n] \backslash S$

Proof.

$$\begin{aligned}
\|\widetilde{x}\|_1 &\geq \|\widetilde{x} + \Delta\|_1 \\
&= \|(\widetilde{x} + \Delta)_S\|_1 + \|(\widetilde{x} + \Delta)_{\overline{S}}\|_1 \\
&= \|\widetilde{x}_S + \Delta_S\|_1 + \|\Delta_{\overline{S}}\|_1 \\
&\geq \|\widetilde{x}\|_1 - \|\Delta_S\|_1 + \|\Delta_{\overline{S}}\|_1
\end{aligned}$$

□

Next we shall show the 2-norm part. To this end, let us order each $i \in \overline{S}$ by its corresponding Δ_i value in nonincreasing order and partition the ordered index set $\overline{S}$ into blocks $B_1, B_2, \ldots$ of size $2r$ each (except possibly for the last block of remaining indices). Namely, B_1 are the indices of the $2r$ largest coordinates of $\Delta_{\overline{S}}$ etc.. This choice implies that for every $i \in B_{j+1}$,

$$|\Delta_i| \leq \frac{\|\Delta_{B_j}\|_1}{2r}$$

Summing over $i \in B_{j+1}$, we have

$$\begin{aligned}
\sum_{i \in B_{j+1}} |\Delta_i|^2 &\leq \frac{\|\Delta_{B_j}\|_1^2}{2r} \\
\implies \|\Delta_{B_{j+1}}\|_2 &\leq \frac{\|\Delta_{B_j}\|_1}{\sqrt{2r}}
\end{aligned}$$

Then we can bound

$$\begin{aligned}
\sum_{j \geq 1} \|\Delta_{B_{j+1}}\|_2 &\leq \sum_{j \geq 1} \frac{\|\Delta_{B_j}\|_1}{\sqrt{2r}} \\
&= \frac{1}{\sqrt{2r}} \|\Delta_{\overline{S}}\|_1 \\
&\leq \frac{1}{\sqrt{2r}} \|\Delta_S\|_1 && \text{(claim 3.4)} \\
&\leq \frac{1}{\sqrt{2}} \|\Delta_S\|_2 && (\|z\|_1 \leq \sqrt{d}\,\|z\|_2 \text{ for } z \in \mathbb{R}^d)
\end{aligned}$$

Now we can derive a contradiction:

$$\begin{aligned}
0 &= \|A\Delta\|_2 \\
&\geq \|A_{S \cup B_1} \Delta_{S \cup B_1}\|_2 - \sum_{j \geq 2} \|A_{B_j} \Delta_{B_j}\|_2 && \text{(triangle inequality)} \\
&\geq (1 - \epsilon_0) \|\Delta_{S \cup B_1}\|_2 - (1 + \epsilon_0) \sum_{j \geq 2} \|\Delta_{B_j}\|_2 && \text{(almost-isometry)} \\
&\geq (1 - \epsilon_0) \|\Delta_S\|_2 - \frac{1 + \epsilon_0}{\sqrt{2}} \|\Delta_S\|_2 && \text{(above bound)} \\
&= \|\Delta_S\|_2 \left(1 - \epsilon_0 - \frac{1 + \epsilon_0}{\sqrt{2}}\right)
\end{aligned}$$

When ϵ_0 is small enough, this is a contradiction. □

3.2 Proof of main theorem

In light of lemma 3.3, we need to prove that if A has JL property as specified in lemma 1.2, then with probability at least $1 - e^{c_1 m}$ for some constant c_1, A is a $3r$-restricted ϵ_0-almost isometry. For notational ease, let $t := 3r$. Note that we only need to consider unit vectors. Below we show an upper bound on the probability that A does not have this property on S^{t-1}, where $S^{t-1} := \{x \in \mathbb{R}^t : \|x\|_2 = 1\}$.

Suppose A does not have this property. Then there exists a t-element set $T \subset [n]$ such that A_T is not an ϵ_0-almost isometry. Note there are at most $\binom{n}{t}$ choices for T. Let A_T be the submatrix of A whose columns are indexed by T and rows the same as A. Below we use a union-bound type argument to upper bound the probability that A_T is not an ϵ_0-almost isometry. If we can do so, since there are at most $\binom{n}{t}$ choices for T, we will be done by a union-bound argument and choosing properly the magnitude of all parameters. To state two lemmas (see lemma 1.1 and 1.3 in Matousek's notes and the proofs therein), we need the following definition:

Definition 3.5. For a given set C in a Euclidean space, a subset $S \subset C$ is said to be ϵ-dense, if $\forall x \in C, \exists x' \in S$ such that $\|x - x'\|_2 \leq \epsilon$.

The lemma below says roughly that the behaviour (in the sense of preserving distance) of F on a dense set is a 'good approximation' of the behaviour on the whole set, and how good the approximation is depends on density of the subset.

Lemma 3.6. *Let $\epsilon \in (0, \frac{1}{3})$, let $N \subset S^{t-1}$ be ϵ-dense. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map satisfying*

$$1 - \epsilon \leq \|F(q)\|_2 \leq 1 + \epsilon, \forall q \in N$$

Then F is a 3ϵ-almost isometry.

Let us fix an $\frac{\epsilon_0}{3}$-net N_T on S^{t-1} (we will show the finiteness of N_T later). Recall A_T is not an ϵ_0-almost isometry, so the above lemma says that there exists $q \in N_T$ such that the following does *not* hold:

$$1 - \frac{\epsilon_0}{3} \leq \|A_T q\| \leq 1 + \frac{\epsilon_0}{3}$$

By JL property, we know that for a fixed unit vector q, the above inequality does not hold with probability at most $e^{-c'm}$ for some positive c' depending on ϵ_0. Now we can use a union bound argument to finish the proof if we can upper bound the size of N_T. The following lemma allows us to do this.

Lemma 3.7. $\forall \epsilon \in (0, 1)$*, there exists an ϵ-dense set $N \subset S^{t-1}$ such that $|N| \leq (\frac{4}{\epsilon})^t$.*

There are at most K^t choices for N_T with K depending on ϵ_0 and $\binom{n}{t}$ choices for $T \subset [n]$. Therefore the probability that A does not have t-restricted ϵ_0-almost isometry property is at most

$$\binom{n}{t} K^t e^{-c'm} \leq (\frac{ne}{t})^t K^t e^{-c'm} = \exp(3r(\ln \frac{ne}{3r}) + \ln K - c_2 m)$$

With the assumption that $m \geq Cr \log(n/r)$ from our main theorem, it is not hard to check that the last expression is bounded above by $e^{-c_2 m/2}$ if C is sufficiently large. So we are done.