CS 270 Algorithms **Spring 2021**

Lecture 10: Compressed sensing, RIP property

Lecturer: Prasad Raghavendra Scribe: Adrian Fan, Paul Shao

10.1 Proof of Oblivious Subspace Embedding

First we will finish the proof of oblivious subspace embedding from last lecture.

Theorem 10.1 (Oblivious Subspace Embedding)**.** *Let S be any subspace of $\mathbb{R}^n$ with dimension d, and let $G = \frac{1}{\sqrt{t}}(N(0,1))^{t\times n}$ be a random Gaussian matrix. If $t = \Omega\left(\dfrac{d+\log(1/\delta)}{\epsilon^2}\right)$, then with probability $1-\delta$, for all $\vec{u} \in S$, $\|G\vec{u}\| \in ((1-\epsilon)\|\vec{u}\|, (1+\epsilon)\|\vec{u}\|)$.*

Proof. First recall the following definition:

Definition 10.2 (γ-net)**.** *A γ-net $\mathcal{N}$ of the ball(0, 1) in $\mathbb{R}^d$, i.e. the ball centered at the origin with radius 1, is a set of points such that for all x in the ball, there exists $y \in \mathcal{N}$ such that $\|x-y\| \le \gamma$.*

Now recall two observations made from last lecture:

Observation 1: There exists a γ-net $\mathcal{N}$ with $|\mathcal{N}| \le \left(\dfrac{1}{\gamma}\right)^d$.

Observation 2: By distributional Johnson–Lindenstrauss. with probability $1-\delta$, for all $x, y \in \mathcal{N}$, $\langle Gx, Gy\rangle \in (\langle x, y\rangle - \epsilon, \langle x, y\rangle + \epsilon)$.

Lemma 10.3. *Consider a γ-net $\mathcal{N}$ with $\gamma \le \frac{1}{2}$. For all $x \in Ball(0,1)$, x can be written as $x = \sum_{i=1}^{\infty} \alpha_i y_i$ where $y_i \in \mathcal{N}$ and $|\alpha_i| \le \dfrac{1}{2^{i-1}}$*

Proof. Let $x_1 = x$ and let y_1 be the point in $\mathcal{N}$ closest to x_1.
Write $x_1 = y_1 + (x_1 - y_1) = y_1 + \|x_1 - y_1\| \cdot \dfrac{x_1 - y_1}{\|x_1 - y_1\|}$.
Note that $y_1 \in \mathcal{N}$ and $\|x_1 - y_1\| \le \delta$. Now set $x_2 = \dfrac{x_1 - y_1}{\|x_1 - y_1\|}$. Let y_2 the the point in $\mathcal{N}$ closest to x_2. Then write $x_1 = y_1 + \|x_1 - y_1\| \cdot \left(y_2 + \|x_2 - y_2\| \cdot \dfrac{x_2 - y_2}{\|x_2 - y_2\|}\right)$
Then set $x_3 = \dfrac{x_2 - y_2}{\|x_2 - y_2\|}$, so $x_1 = y_1 + \|x_1 - y_1\| x_2 + \|x_1 - y_1\|\|x_2 - y_2\| y_3$. Note that $\|x_1 - y_1\|, \|x_2 - y_2\| \le \gamma$.

Continuing in this fashion, we obtain an infinite series of the form $x = \sum_{i=1}^{\infty} \alpha_i y_i$ where $\alpha_i \leq \gamma^{n-1}$. If we pick $\gamma \leq \frac{1}{2}$, we obtain the desired result. $\square$

Now we will prove the Oblivious Subspace Embedding theorem. Consider any $x \in \mathrm{Ball}(0,1)$. Then

$$\begin{aligned}
\|Gx\|^2 &= \langle Gx, Gx \rangle \\
&= \left\langle G\left(\sum_{i=1}^{\infty} \alpha_i y_i\right), G\left(\sum_{i=1}^{\infty} \alpha_i y_i\right)\right\rangle \\
&= \sum_{i,j} \alpha_i \alpha_j \langle Gy_i, Gy_j \rangle \\
&= \sum_{i,j} \alpha_i \alpha_j (\langle y_i, y_j \rangle \pm \epsilon) \\
&= \left\langle \sum_{i=1}^{\infty} \alpha_i y_i, \sum_{j=1}^{\infty} \alpha_j y_j \right\rangle \pm \epsilon \cdot \left(\sum_{i,j} \alpha_i \alpha_j\right) \\
&= \langle x, x \rangle + \epsilon \cdot \left(\sum_{i=1}^{\infty} \alpha_i\right)^2 \\
&= \|x\|^2 + O(\epsilon)
\end{aligned}$$

where the O bound follows from the fact that $\sum_{i=1}^{\infty} \alpha_i$ converges.

$\square$

10.2 Compressed Sensing

Compressed sensing is an application of dimension reduction for sparse vectors. As a real life application, images and audio are often sparse in some basis. This fact underlies JPEG compression: a typical image is not a random vector, so it is representable by a sparse vector. In our application we have a signal x in n dimensions but in some basis, most of the n coordinates of x are approximately 0.

Definition 10.4 (k-sparse vector)**.** *A vector $x \in \mathbb{R}^n$ is k-sparse if $< k$ coordinates of x are nonzero. Similarly, x is ϵ-approximately k-sparse if there exists $S \subseteq [n] = \{1, 2, \ldots, n\}$ where $|S| = k$ and $\|x_{\bar{S}}\| \leq \epsilon \|x\|$, (where x_S denotes the vector formed by taking x and replacing all coordinates not in S with 0).*

In this lecture, we will only deal with exactly k-sparse vectors, not approximately k-sparse vectors. The setup of compressed sensing is as follows: we are given a signal $x = (x_1, \ldots, x_n)$ that is k-sparse. We want to use $\ll n$ measurements to recover x. We have a measurement matrix $M \in \mathbb{R}^{t \times n}$, where $t \approx k \log n$, and we measure the coordinates of Mx, namely $\langle M_1, x \rangle, \ldots, \langle M_t, x \rangle$. Then given the system $Mx = b$, we want to recover a k-sparse solution x to the system. We have $k \log n$ equations in n variables, but we want to be able to uniquely recover x given the extra information that x is k-sparse.

Intuitively, the measurements $\langle M_1, x\rangle, \ldots, \langle M_t, x\rangle$ could be the measurements made by a camera. The camera won't measure x; it instead directly observes the coordinates of Mx. This problem is difficult because we don't know which coordinates of x are nonzero. We won't be able to measure all coordinates of x with a few measurements, but we still want to recover a sparse vector. Compressed sensing has applications in MRI, where it is used to reduce the number of images required.

To summarize, our input is a system $Mx = b$, and we want to return a k-sparse solution to the system. This problem can be rephrased as an optimization problem:

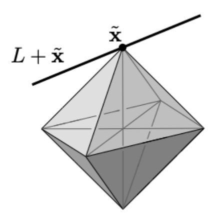
$$\begin{gathered}\min \|x\|_0 \\ \text{subject to } Mx = b\end{gathered}$$

where $\|x\|_0$ denotes the number of nonzero entries of x. However, this optimization problem is not so nice because $\|x\|_0$ is not a convex or smooth function. Often, we can replace $\|x\|_0$ with $\|x\|_1$, so our optimization problem becomes

$$\begin{gathered}\min \|x\|_1 = \sum_i |x_i| \\ \text{subject to } Mx = b\end{gathered}$$

It turns out that minimizing the l_1 norm $\|x\|_1$ typically gives sparse vectors. For a large fraction of hyperplanes M, minimizing the l_1 norm works.

Here's a geometric illustration on why that is the case:



As we can see, the l_1 ball $\{x : \|x\|_1 \leq 1\}$ in $\mathbb{R}^3$ is a regular octahedron containing lower-dimensional facets. A feasible solution to the linear program under l_1 constraint can usually be viewed as the line (or plane) representing the linear constraint touching the l_1 ball at exactly one of its vertices.

It can be proven that minimizing the l_1 norm also works if M is a random matrix, such as a Gaussian matrix or a matrix with random ± 1 entries. Now we introduce some terminology:

Definition 10.5 (Restricted Isometry Property (RIP))**.** *A matrix $M \in \mathbb{R}^{t \times n}$ satisfies (k, ϵ) Restricted Isometry Property (RIP) if for every k-sparse x, $\|Mx\| \in ((1-\epsilon)\|x\|, (1+\epsilon)\|x\|)$.*

Note that the set of all k-sparse vectors is

$$\bigcup_{S \subseteq [n], |S| = k} \{x : x_i = 0 \text{ for all } i \notin S\}$$

which is a union of $\binom{n}{k}$ subspaces. For any subspace, we want the probability of not preserving distances to be less than $\delta/\binom{n}{k}$ so we can use Union Bound. Then by using Oblivious Subspace Embedding, if we pick $t = \dfrac{k + \log\binom{n}{k} + \log(1/\delta)}{\epsilon^2} = O\left(\dfrac{k + \log\binom{n}{k}}{\epsilon^2}\right)$, with probability $1-\delta$, M satisfies (k,ϵ) RIP.

We have shown that random Gaussian matrices usually have the RIP property, but we don't have an explicit construction for a matrix with the RIP property. We can't be sure a specific random Gaussian matrix will work; there is still a probability of error. In addition, for most applications, choosing M to have random ± 1 entries works as well as choosing M to have Gaussian entries. Finally, there is no known efficient algorithm for verifying the RIP property.

Now we will examine why RIP matrices are useful for compressed sensing:

Lemma 10.6. *Suppose M satisfies $(3k, 0.01)$ RIP. Then for any k-sparse $x \in \mathbb{R}^n$, basis pursuit or ℓ_1 norm minimization recovers x exactly from Mx.*

Proof. Consider the following optimization problem

$$x' = \arg\min_{Mx=b} ||x||_1$$
$$x \text{ is k-sparse}$$

By contradiction, suppose there exists a different solution x' with smaller or equal l_1 norm. Let $\Delta = x' - x$, we have the following:

$$M\Delta = M(x'-x) = Mx' - Mx = b - b = 0$$
$$||x'||_1 = ||x+\Delta||_1 \leq ||x||_1$$
$$\text{Assume } ||\Delta||_1 = 1$$

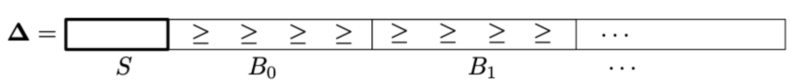
Let $S = \{i : x_i \neq 0\}$, $|S| \leq k$, we observe that at least half of the l_1 norm of Δ has to live on S:

$$||\Delta_S||_1 \geq ||\Delta_{\overline{S}}||_1$$

where ΔS denotes the vector consisting of the components of Δ indexed by S, and $\overline{S} = \{1, 2, \ldots, n\} \setminus S$. Indeed, when Δ is added to x, its components outside S only increase the l_1 norm, and since $||x+\Delta||_1 \leq ||x||_1$, the components in S must at least compensate for this increase.

Since $||\Delta||_1 = 1$, we have $||\Delta S||_1 \geq \frac{1}{2}$.

Let $B_0 \supset \overline{S}$ consist of the indices of the $2r$ largest components of $\Delta_{\overline{S}}$,B_1 are the indices of the next $2r$ largest components, and so on (the last block may be smaller).



We have $\|M_{S\cup B_0}\|_2 \geq (1-\epsilon)\,|\Delta_{S\cup B_0}|_2 \geq (1-\epsilon)\|\Delta_S\|_2 \geq \frac{1}{\sqrt{2k}}(1-\epsilon)$

In addition, since $\|\Delta_S\|_1 \geq \frac{1}{2}$, $\|\Delta_S\|_2 \geq \frac{1}{\sqrt{2k}}$, and therefore $\|\Delta_{\overline{S}}\|_2 \leq \frac{1}{2}$.

now, for any index i in the chunk:

$$B_i \leq \frac{\|\Delta_{B_i-1}\|_1}{2k}$$

Bounding with the following lemma about metric norms,

$$\text{For } v \in \mathbb{R}^d,\ \|v\|_2 \geq \frac{\|v\|_1}{\sqrt{d}},$$

we have:

$$\begin{aligned}
\|\Delta_{B_i}\|_2 &\leq \left(2k\left(\frac{\|\Delta_{B_i-1}\|_1}{2k}\right)^2\right)^{1/2} \\
&\leq \frac{\|\Delta_{B_i-1}\|_1}{\sqrt{2k}}
\end{aligned}$$

Summing over all indices, this gives us:

$$\begin{aligned}
\sum_{i=1}^{\infty} \|\Delta_{B_i}\|_2 &\leq \sum_{i=1}^{\infty} \frac{\|\Delta_{B_i-1}\|_1}{\sqrt{2k}} \\
&= \frac{\|\Delta_{\overline{S}}\|_1}{\sqrt{2k}} \\
&= \frac{1}{2\sqrt{2k}} \\
&\approx \frac{0.4}{\sqrt{k}}
\end{aligned}$$

Hence, we have proven the following claim:

$$\sum_{j\geq 1} \|\Delta_{B_j}\|_2 \leq \frac{0.4}{\sqrt{k}}$$

However, since we assumed the following:

$$M\Delta = \underbrace{M\Delta_{S\cup B_0}}_{\geq \frac{1}{\sqrt{2k(1-\epsilon)}}} + \underbrace{\sum_{j\geq 1} \|\Delta_{B_j}\|_2}_{\leq \frac{0.4}{\sqrt{k}}(1+\epsilon)}$$

The first part of the sum $M\Delta_{S\cup B_0}$ is clearly much larger than the second part $\sum_{j\geq 1} \|\Delta_{B_j}\|_2$, hence, it is impossible for $M\Delta$ to be 0. Therefore, we've reached a contradiction with our original claim that $M\Delta = 0$. $\square$