Scribed by Arun Ganesh

Lecture 16

In which we show how to use linear programs to find planted sparse vectors in a random subspace.

1 Review: Sparse Reconstruction Problem

Recall the planted sparse vector problem introduced in the last lecture. In the problem, we are given a subspace $S = \text{span}(\mathbf{v}_0, \mathbf{v}_1, \dots \mathbf{v}_d) \subset \mathbb{R}^n$ where $\mathbf{v}_0$ is a vector with at most k non-zero entries, and $\mathbf{v}_1 \dots \mathbf{v}_d$ are random vectors (i.e. their elements are independently drawn from $N(0,1)$). Our goal is to find a k-sparse vector, i.e. one with at most k non-zero entries - ideally, we would like to find $\mathbf{v}_0$.

1.1 Certifying a lack of sparse vectors

In the last lecture, we showed that given the space $S = \text{span}(\mathbf{v}_1, \dots \mathbf{v}_d)$ which does not include the planted sparse vector $\mathbf{v}_0$, we can certify with high probability that no $O(\frac{n}{\sqrt{d}})$-sparse vector exists. Some of the techniques used will be useful finding sparse vectors, so we review the proof here.

We can use the following formulation for finding the sparse vector (recall that $\|\mathbf{v}\|_0$ is the number of non-zero elements of $\mathbf{v}$):

$$\begin{aligned} \min\ & \|\mathbf{z}\|_0 \\ \text{s.t.}\ & \mathbf{z} \in S \\ & \mathbf{z} \neq \mathbf{0} \end{aligned} \tag{1}$$

Let $\mathbf{z}^*$ be an optimal solution to (1), with $\|\mathbf{z}\|_0 = k$. Then $\|\mathbf{z}\|_1 \leq k\|\mathbf{z}\|_\infty$. Thus the optimum solution of (1) is lower bounded by the optimal solution to:

$$\begin{aligned} \min\ & \frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_\infty} \\ \text{s.t.}\ & \mathbf{z} \in S \\ & \mathbf{z} \neq \mathbf{0} \end{aligned} \tag{2}$$

Since $\mathbf{z}^*$ is feasible in (2) and achieves objective k. Similarly, let $\mathbf{z}^*$ be an optimal solution to (2) for which $\frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_\infty} = k$. Let i_{max} be the index of the element of $\mathbf{z}^*$ with the largest absolute value. Then $\mathbf{y} = \frac{1}{\mathbf{z}(i_{max})}\mathbf{z}^*$ satisfies $\mathbf{y} \in S$, $\|\mathbf{y}\|_\infty = 1$, and $\|\mathbf{y}\|_1 = k$. Then, the optimal solution of (2) is lower bounded by:

$$\min_i OPT_i \tag{3}$$

Where:

$$OPT_i = \begin{bmatrix} \min \ \|\mathbf{z}\|_1 \\ \text{s.t.} \ \mathbf{z} \in S \\ \quad \mathbf{z}(i) = 1 \\ \quad \mathbf{z} \neq \mathbf{0} \end{bmatrix}$$

Since $\mathbf{y}$ gives a solution achieving objective k to the program defined by $OPT_{i_{max}}$. Note that for each i, the program defined by OPT_i can be written as a linear program. Therefore, we can efficiently solve for the value of (3) and thus obtain a lower bound certificate for the sparse vector problem in a random subspace.

Let A be the matrix defined as follows:

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_d \\ | & | & & | \end{bmatrix}$$

i.e, A is the matrix whose columns are the basis vectors of S. Then $S = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^d\}$. Then for a fixed i:

$$OPT_i = \begin{bmatrix} \min \ \|A\mathbf{x}\|_1 \\ \text{s.t.} \ \mathbf{x} \in \mathbb{R}^d \\ \quad (A\mathbf{x})(i) = 1 \\ \quad \mathbf{x} \neq \mathbf{0} \end{bmatrix}$$

We can use the following theorem (which we state without proof) to lower bound OPT_i with high probability:

Theorem 1 *If $A \in \mathbb{R}^{n \times d}$ where $d \leq \frac{n}{2}$ and the elements of A are sampled from $N(0,1)$ independently, then with high probability:*

$$\forall \mathbf{x} \in \mathbb{R}^d, \|A\mathbf{x}\|_1 \geq \Omega(\sqrt{n})\|A\mathbf{x}\|_2$$

Then with high probability OPT_i can be lower bounded by:

$$\begin{aligned} &\min \ \Omega(\sqrt{n})\|A\mathbf{x}\|_2 \\ &\text{s.t.} \ \mathbf{x} \in \mathbb{R}^d \\ &\quad (A\mathbf{x})(i) = 1 \\ &\quad \mathbf{x} \neq \mathbf{0} \end{aligned} \tag{4}$$

We can use the following theorem (whose proof we also omit) to lower bound (4):

Theorem 2 *If $A \in \mathbb{R}^{n\times d}$ and the elements of A are sampled from $N(0,1)$ independently, then with high probability:*

$$A^T A \succeq \frac{n}{2} I$$

Then by Theorem 2, with high probability:

$$\|A\mathbf{x}\|_2^2 = \mathbf{x}^T A^T A x \geq \frac{n}{2}\mathbf{x}^T I \mathbf{x} = \frac{n}{2}\|x\|_2^2 \tag{5}$$

Taking the square root of the first and last term in (5) gives $\|A\mathbf{x}\|_2 \geq \Omega(\sqrt{n})\|bx\|_2$, and thus we can lower bound (4) by:

$$\begin{aligned} \min\ & \Omega(n)\|\mathbf{x}\|_2 \\ \text{s.t.}\ & \mathbf{x} \in \mathbb{R}^d \\ & (A\mathbf{x})(i) = 1 \\ & \mathbf{x} \neq \mathbf{0} \end{aligned} \tag{6}$$

Note that $(A\mathbf{x})(i) = \langle \mathbf{x}, (A_{i,1}, A_{i,2} \dots A_{i,d})\rangle$. By applying the Cauchy-Schwarz inequality, we get that $(A\mathbf{x})(i) \leq \|\mathbf{x}\|_2 \cdot \|(A_{i,1}, A_{i,2} \dots A_{i,d})\|_2$. This lets us lower bound 6 by:

$$\begin{aligned} \min\ & \Omega(n)\|\mathbf{x}\|_2 \\ \text{s.t.}\ & \mathbf{x} \in \mathbb{R}^d \\ & \|\mathbf{x}\|_2 \cdot \|(A_{i,1}, A_{i,2} \dots A_{i,d})\|_2 \geq 1 \\ & \mathbf{x} \neq \mathbf{0} \end{aligned} \tag{7}$$

Then, with high probability $\|(A_{i,1}, A_{i,2} \dots A_{i,d})\|_2 \leq O(\sqrt{d})$. Thus for any feasible $\mathbf{x}$ in (7), $\|\mathbf{x}\|_2 \geq \Omega(\frac{1}{\sqrt{d}})$, which means (7) is at least $\Omega(\frac{n}{\sqrt{d}})$ with high probability. By applying union bound, we get that for all i, this holds with high probability. The minimum value of (7) across all i is a lower bound to the sparsity of the sparsest vector in S, so we conclude that with high probability each vector in S has sparsity at least $\Omega(\frac{n}{\sqrt{d}})$.

2 Finding Planted Sparse Vectors

Thus far, we have only shown that we can certify a lower bound for the sparsest vector in a random subspace. Now, we want to argue that we can easily find a planted vector $\mathbf{v}_0$ with at most $\|\mathbf{v}_0\|_0 = k \leq c\frac{n}{\sqrt{d\log n}}$ non-zero elements (for some constant c).

In particular, we will argue that with high probability the following is minimized by a multiple of $\mathbf{v}_0$:

$$\min_i \begin{bmatrix} \min & \|\mathbf{z}\|_1 \\ \text{s.t.} & \mathbf{z} \in S \\ & \mathbf{z}(i) = 1 \\ & \mathbf{z} \neq \mathbf{0} \end{bmatrix} \tag{8}$$

Since (8) and its minimizer can be computed by solving n linear programs, we can find $\mathbf{v}_0$ in polynomial time.

Without loss of generality, assume that:

- $\mathbf{v}_0(1)$ is the largest element of $\mathbf{v}_0$
- $\mathbf{v}_0(1) = 1$
- $\mathbf{v}_0(1), \mathbf{v}_0(2), \ldots \mathbf{v}_0(k)$ are the non-zero elements of $\mathbf{v}_0$.

Then, define A as follows:

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_0 & \mathbf{v}_1 & \ldots & \mathbf{v}_d \\ | & | & & | \end{bmatrix}$$

It suffices to prove that

$$\begin{aligned} \min \ & \|A\mathbf{x}\|_1 \\ \text{s.t.} \ & \mathbf{x} \in \mathbb{R}^{d+1} \\ & (A\mathbf{x})(1) = 1 \\ & \mathbf{x} \neq \mathbf{0} \end{aligned} \tag{9}$$

Is uniquely minimized by $\mathbf{x} = (1, 0, 0, \ldots 0)$.

2.1 Proof strategy

Before starting the proof, we go over the high level strategy we want to use. Note that $\mathbf{v}_0$ looks like:

$$\begin{bmatrix} 1 \\ [-1, 1] \\ [-1, 1] \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

That is, the first element is 1, the next $k-1$ elements are in $[-1, 1]$, and then all remaining elements are 0.

Suppose we start with the solution $\mathbf{x} = (1, 0, 0, \ldots 0)$ to (9) (say that $\mathbf{x}$ is indexed from 0 to d, so that $A\mathbf{x} = \sum_i \mathbf{x}(i)\mathbf{v}_i$). Then, the last $n-k$ elements of $A\mathbf{x}$ are all zero. Consider modifying this solution. If we make any of $\mathbf{x}(1), \ldots \mathbf{x}(d)$ non-zero, then the last $n-k$ elements of $A\mathbf{x}$ can only increase in absolute value, and thus their contribution to $\|A\mathbf{x}\|_1$ will increase. Then, it only makes sense to make any of $\mathbf{x}(1), \ldots \mathbf{x}(d)$ non-zero if doing so decreases the contribution of the first k elements of $A\mathbf{x}$ to $\|A\mathbf{x}\|_1$ by a larger amount than the contribution of the last $n-k$ elements is increased by. However, since $\mathbf{v}_1, \mathbf{v}_2 \ldots$ are chosen randomly, and $n-k \gg k$, it should be unlikely that this is possible.

2.2 Proof of unique optimum

Now to formalize the proof. Split A into submatrices as follows:

$$A = \begin{bmatrix} \mathbf{v}'_0 & M_1 \\ \mathbf{0} & M_2 \end{bmatrix}$$

Where $\mathbf{v}'_0$ is the vector consisting of the non-zero elements of $\mathbf{v}_0$, $\mathbf{0}$ is the all-zeroes vector, M_1 is the matrix whose columns are the first k elements of $\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_d$, and M_2 is the matrix whose columns are the last $n-k$ elements of $\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_d$. Consider a solution $\mathbf{x}$, written as:

$$\mathbf{x} = \begin{bmatrix} x_0 \\ \mathbf{x}' \end{bmatrix}$$

Where $\mathbf{x}'$ consists of the last d elements of $\mathbf{x}$. Then:

$$\|A\mathbf{x}\|_1 = \left\| x_0\mathbf{v}_0 + \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \mathbf{x}' \right\|_1 =$$
$$\left\| x_0\mathbf{v}_0 + \begin{bmatrix} M_1 \\ \mathbf{0} \end{bmatrix} \mathbf{x}' + \begin{bmatrix} \mathbf{0} \\ M_2 \end{bmatrix} \mathbf{x}' \right\|_1 \tag{10}$$

Where $\mathbf{0}$ is the all-zeroes matrix. Note that the non-zero elements of the first and second term in 10 and the non-zero elements of the third term are in disjoint positions. Then 10 is equal to:

$$\left\| x_0\mathbf{v}_0 + \begin{bmatrix} M_1 \\ \mathbf{0} \end{bmatrix} \mathbf{x}' \right\|_1 + \left\| \begin{bmatrix} \mathbf{0} \\ M_2 \end{bmatrix} \mathbf{x}' \right\|_1 \tag{11}$$

By triangle inequality, we get that 11 is lower bounded by:

$$\|x_0\mathbf{v}_0\|_1 - \left\| \begin{bmatrix} M_1 \\ \mathbf{0} \end{bmatrix} \mathbf{x}' \right\|_1 + \left\| \begin{bmatrix} \mathbf{0} \\ M_2 \end{bmatrix} \mathbf{x}' \right\|_1 \tag{12}$$

Since x_0 is a scalar and the $\mathbf{0}$ portion of the second and third term in 12 contribute nothing to the norm, we can rewrite 12 as:

$$x_0 \left\|\mathbf{v}_0\right\|_1 - \left\|M_1\mathbf{x}'\right\|_1 + \left\|M_2\mathbf{x}'\right\|_1 \tag{13}$$

We now use two lemmas to lower bound (13).

Lemma 3 *If M is a $k \times d$ matrix whose elements are sampled from $N(0,1)$ independently, with high probability:*

$$\forall \mathbf{x} \in \mathbb{R}^d : \|M\mathbf{x}\|_1 \le 2k\|\mathbf{x}\|_1$$

Proof: Let $\|M\|_{1\to1} = \max_{\mathbf{x}\neq 0} \frac{\|M\mathbf{x}\|_1}{\|\mathbf{x}\|_1}$. We are trying to show that $\|M\|_{1\to1} \le 2k$. Without proof, we claim that:

$$\|M\|_{1\to1} = \max_i \|(M_{1,i}, M_{2,i} \dots M_{k,i})\|_1$$

Fix some i. Then:

$$\|(M_{1,i}, M_{2,i} \dots M_{k,i})\|_1 = \sum_{j=1}^{k} |M_{j,i}| = \sum_{j=1}^{k} \text{sign}(M_{j,i})M_{j,i}$$

We want to show this is at most $2k$ with high probability. We show a stronger statement - consider fixing k bits $b_1, \dots b_k$ where $b_j \in \{-1, 1\}$. Then $\sum_{j=1}^{k} b_j M_{j,i}$ is a sum of k $N(0,1)$ random variables, and is thus a $N(0,k)$ random variable. We know by properties of Gaussians that:

$$Pr[N(0,k) \ge t] \le e^{-t^2/2k}$$

In particular, if $t = 2k$ then we get:

$$Pr[N(0,k) \ge 2k] \le e^{-2k}$$

Thus for any fixed choice of bits $b_1, \dots b_k$, $\sum_{j=1}^{k} b_j M_{j,i}$ is at least $2k$ with probability at most e^{-2k}. Taking the union over all choices of bits, we get that for all choices of bits, $\sum_{j=1}^{k} b_j M_{j,i} \le 2k$ and thus $\|(M_{1,i}, M_{2,i} \dots M_{k,i})\|_1 \le 2k$ with probability at most $2^k e^{-2k}$ (since $\|(M_{1,i}, M_{2,i} \dots M_{k,i})\|_1$ is equal to $\sum_{j=1}^{k} b_j M_{j,i}$ for $b_j = \text{sign}(M_{j,i})$). Taking the union over all choices of i, we get that $\|M\|_{1\to1} \ge 2k$ with probability at most $d2^k e^{-2k}$, which vanishes quickly with k. $\square$

Lemma 4 *If M is a $n \times d$ matrix whose elements are sampled from $N(0,1)$ independently and $d \leq \frac{n}{16}$, with high probability:*

$$\forall \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq 0 : \|M\mathbf{x}\|_1 \geq \Omega(\frac{n}{\sqrt{d}})\|\mathbf{x}\|_1$$

PROOF: By applying Theorem 1 and Theorem 2:

$$\|M\mathbf{x}\|_1 \geq \Omega(\sqrt{n})\|M\mathbf{x}\|_2 \geq \Omega(n)\|\mathbf{x}\|_2 \geq \Omega(\frac{n}{\sqrt{d}})\|\mathbf{x}\|_1$$

□

Using Lemmas 3 and 4 and the fact that (13) lowerbounds $\|A\mathbf{x}\|_1$, we get that:

$$\|A\mathbf{x}\|_1 \geq |x_0|\|\mathbf{v}_0\|_1 - 2k\|\mathbf{x}'\|_1 + \Omega(\frac{n-k}{\sqrt{d}})\|\mathbf{x}'\|_1$$

If $\mathbf{x}$ is the optimal solution, then $\|\mathbf{v}_0\|_1 \geq \|A\mathbf{x}\|_1$ since $\mathbf{v}_0$ is a feasible solution. Assuming $k \leq c\frac{n}{\sqrt{d}\sqrt{\log n}}$ for some sufficiently small c we get:

$$\|\mathbf{v}_0\|_1 \geq |x_0| \cdot \|\mathbf{v}_0\|_1 + 4k\sqrt{\log n} \cdot \|\mathbf{x}'\|_1 \tag{14}$$

Let us call $\mathbf{a} = (M_{1,1}, \dots, M_{1,d})$ the k-dimensional vector corresponding to the first entries of $\mathbf{v}_1, \dots, \mathbf{v}_d$. We have the constraint

$$x_0 + \langle \mathbf{x}', \mathbf{a} \rangle = 1$$

By properties of Gaussians, with high probability:

$$\|\mathbf{a}\|_\infty \leq 2\sqrt{\log n}$$

Combining these two statements with (14) we get:

$$\begin{aligned}
\|\mathbf{v}_0\|_1 &\geq |1 - \langle \mathbf{x}', \mathbf{a} \rangle| \cdot \|\mathbf{v}_0\|_1 + 2k \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1 \\
&\geq (1 - \langle \mathbf{x}', \mathbf{a} \rangle) \cdot \|\mathbf{v}_0\|_1 + 2k \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1 \\
&\geq (1 - \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1) \cdot \|\mathbf{v}_0\|_1 + 2k \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1
\end{aligned}$$

Where the last step is because $\langle \mathbf{x}', \mathbf{a} \rangle = \sum_j \mathbf{x}'(j)\mathbf{a}(j) \leq \sum_j \mathbf{x}'(j) \max_k \mathbf{a}(k)$. Rearranging terms gives:

$$\|\mathbf{v}_0\|_1 \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1 \geq 2k \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1 \geq 2\|\mathbf{v}_0\|_1 \cdot \|\mathbf{a}\|_\infty \cdot \|\mathbf{x}'\|_1$$

$$\|\mathbf{x}'\|_1 \geq 2\|\mathbf{x}'\|_1$$

which implies $\mathbf{x}' = \mathbf{0}$.