

Lecture 0: Guest Lecture by Venkat Guruswami:
Euclidean Subspaces and Compressed SensingLecturer: *Sanjeev Arora*Scribe: *Moritz Hardt*

Overview

In this lecture we study high-dimensional subspaces $X \subseteq \mathbf{R}^N$ in which we have $\|x\|_1 \approx \sqrt{N} \|x\|_2$ for all vectors $x \in X$. We discuss a recent result that shows how to construct such subspaces explicitly building on the theory of Expander Codes. Along the way, we recall some notions from metric spaces and metric embeddings. Finally, we learn about an application called Compressed Sensing.

1 Almost Euclidean Subspaces of ℓ_1^N

The metric space ℓ_p^N is defined as the N -dimensional real space \mathbf{R}^N endowed with the ℓ_p -norm. At this point, we recall the definition of the ℓ_p -norm of a vector $x \in \mathbf{R}^N$.

$$\|x\|_p := (\sum_i |x_i|^p)^{1/p} \quad (\ell_p\text{-norm})$$

If $p = \infty$, we let $\|x\|_\infty = \max_i |x_i|$.

We are interested in high-dimensional subspaces of ℓ_1^N that are “similar” to the Euclidean space in the following sense. We say a subspace $X \subseteq \mathbf{R}^N$ is an *almost Euclidean* subspace of ℓ_1^N , if

$$\text{for all } x \in X, \text{ we have } \|x\|_1 \geq \Omega(\sqrt{N}) \|x\|_2. \quad (1)$$

To understand this property, we note that generally

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{N} \|x\|_2, \quad (2)$$

where the second inequality follows from Cauchy-Schwartz. Furthermore, the Cauchy-Schwartz inequality holds with equality for all vectors x in the *one*-dimensional subspace spanned by the all-ones vector. Therefore, it is not difficult to achieve Property 1 when the subspace X has small dimension. But we want the dimension to be very large, let us say, $\dim(X) = \Omega(N)$. As it turns out, even with this additional requirement, such subspaces exist. In fact, in a certain sense most subspaces of dimension, say, $N/2$ satisfy the desired property.

THEOREM 1 (KASHIN, FIGIEL-LINDENSTRAUSS '77)

A random subspace $X \subseteq \mathbf{R}^N$ of dimension $N/2$ is almost Euclidean.

How do we choose the random subspace? For instance, we can define it as the kernel of a random $N/2 \times N$ sign matrix with independent and identically distributed ± 1 entries or Gaussian entries.

1.1 Connection to Metric Embeddings

Let A and B be two metric spaces. A function $E : A \rightarrow B$ is called a metric embedding. Suppose, we have for all $x \in A$,

$$\frac{1}{\Gamma} \|x\|_A \leq \|E(x)\|_B \leq \|x\|_A.$$

Then, the quantity Γ in the first inequality is called the *distortion* of the embedding E . As a remark, such embedding are called *contractive*, since norms in A do not expand in B , but they might shrink.

How does this notion relate to our previous discussion? Suppose we have a d -dimensional almost Euclidean subspace $X \subseteq \mathbf{R}^N$. Write $X = \{Gy \mid y \in \mathbf{R}^d\}$ as the span of the rows of some matrix G . We will assume without loss of generality that G is orthonormal. Define the mapping $E: \ell_2^d \rightarrow \ell_1^N$ by $E(y) = \frac{1}{\sqrt{N}}Gy$. We have

$$\|E(y)\|_1 = \left\| \frac{Gy}{\sqrt{N}} \right\|_1 \geq \Omega(\sqrt{N}) \left\| \frac{Gy}{\sqrt{N}} \right\|_2 = \Omega(1) \|Gy\|_2 = \Omega(1) \|y\|_2,$$

where the second inequality follows from Property 1 and the last equality holds because G is a rotation. The point is that an almost Euclidean subspace X induces a metric embedding from ℓ_2 into ℓ_1 of constant distortion. This motivates the following definition.

DEFINITION 1 (DISTORTION) *We define the distortion of a subspace $X \subseteq \mathbf{R}^N$ by*

$$\Delta(X) := \sup_{x \in X \setminus \{0\}} \frac{\sqrt{N} \|x\|_2}{\|x\|_1}.$$

In particular, we always have $1 \leq \Delta(X) \leq \sqrt{N}$.

With this definition at hand, let us return to our random matrix model. We denote by $A_{k,N}$ a random $k \times N$ sign matrix, i.e., a matrix with i.i.d. Bernoulli ± 1 entries.

THEOREM 2 (KASHIN '77, GARNAEV-GLUSKIN '84)

Let $X = \ker(A_{k,N})$. Then with high probability X has distortion

$$\Delta(X) \leq \sqrt{\frac{N}{k} \cdot \log \frac{N}{k}}.$$

2 Explicit Constructions

So far we have only considered *random* subspaces. Often we are interested in a stronger result that shows how to *explicitly* construct such subspaces. A construction algorithm is *explicit*, if it runs in deterministic time polynomial in N and outputs the basis vectors of a subspace with the desired properties.

Results due to Rudin '60 and later Linial, London and Rabinovich '95 give simple explicit construction of subspaces with $\dim(X) = \sqrt{N}$ and constant distortion $\Delta(X) = O(1)$. These constructions were improved by Indyk '00 and '07 [Ind07] to

$$\begin{aligned} \dim(X) &= \frac{N}{2^{(\log \log N)^{O(1)}}, \\ \Delta(X) &= 1 + o(1). \end{aligned}$$

However, in both cases $\dim(X) = o(N)$. For $\dim(X) = \Omega(N)$ there was until recently only one Folklore result known that gave a subspace of distortion $N^{1/4}$. This was improved by Guruswami, Lee and Razborov [GLR08] who achieve distortion

$$\Delta(X) = (\log N)^{O(\log \log \log N)}.$$

A different line of research aims at what we call a partial derandomization of Theorem 2. The theorem as is trivially implies a construction algorithm that uses N^2 random bits (pick a random matrix). Artstein-Avidan and Milman '06 give an algorithm with only $O(N \log N)$ random bits. Lovett and Sodin '07 further improve this to $O(N)$ random bits.

2.1 Connection to Expander Codes

Intuitively, a space $X \subseteq \mathbf{R}^N$ has low distortion, if for any $x \in X$ the mass of x is spread over many coordinates. Conversely, if the mass of some $x \in X$ is tightly concentrated on few coordinates, we expect X to have high distortion. In fact, this intuition is correct and it is not difficult to make precise. However, we will not need the details here. Instead we move on to describe an interesting connection between the spreading properties of a subspace and Expander Codes. This is the main technique behind the result of Guruswami, Lee and Razborov [GLR08].

Consider the following general construction of a subspace $X \subseteq \mathbf{R}^N$. Given a bipartite graph $G = (V_L, V_R, E)$ with $V_L = \{1, 2, \dots, N\}$ such that every vertex in V_R has degree d , and given a subspace $W \subseteq \mathbf{R}^d$, we define the subspace $X = X(G, W) \subseteq \mathbf{R}^N$ by

$$X(G, W) = \{x \in \mathbf{R}^N \mid x_{\delta(j)} \in W \text{ for every } j \in V_R\}.$$

Here, $x_{\delta(j)}$ denotes the restriction of the vector x to those coordinates $i \in V_L$ such that $(i, j) \in E$.

In other words, any d -regular bipartite graph allows us to construct a high-dimensional subspace $X \subseteq \mathbf{R}^N$ from a lower dimensional subspace $W \subseteq \mathbf{R}^d$. How is this useful? As it turns out, we can relate the spreading properties of $X(G, W)$ to the expansion of the bipartite graph. Specifically, if W is a subspace with nontrivial distortion (say, $N^{1/4}$) and G is a graph with good expansion properties, then it can be shown [GLR08] that $X(G, W)$ has low-distortion.

As a remark, the above construction has previously been used to construct error correcting codes. Here, the vertices on the left side of the bipartite graph are identified with the bits of the code word. The subspace W is chosen as the $(d - 1)$ -dimensional subspace $\{w \in \mathbf{R}^d \mid \sum_i w_i = 0\}$. Tanner '81 first analyzed this construction. Later, Sipser and Spielman '96 related the quality of such codes to the expansion properties of the bipartite graph.

3 Compressed Sensing

Compressed Sensing is a problem in the area of image and signal processing. The goal is to reconstruct a signal from fewer measurements than what was traditionally done. To be more precise, a signal is a vector $x \in \mathbf{R}^N$ and we are allowed to capture linear measurements

$\langle y, x \rangle$ of x where $y \in \mathbf{R}^N$. To make *few* measurements means that we have some $k \times N$ matrix A the rows of which correspond to measurements. Hence, the information about x that we extract from a measurement is given by a vector $w = Ax$. We want to be able to reconstruct x efficiently. Since this is not in general possible, we furthermore need the promise that x is *sparse* in the sense that $\|x\|_0 = |\text{supp}(x)|$ is small. This leads to the following reconstruction problem.

$$\text{Find } y \text{ such that } \|y\|_0 \text{ is minimized subject to } Ay = w. \quad (3)$$

However, this optimization problem is non-convex and we cannot solve it efficiently in general. Instead we will study the case where it suffices to solve the following relaxation.

$$\text{Find } y \text{ such that } \|y\|_1 \text{ is minimized subject to } Ay = w. \quad (4)$$

Clearly, this problem can be solved by linear programming. How is it related to the distortion of subspaces? Suppose we fix a $k \times N$ matrix A and let $X = \ker(A)$. Then, we can show the following theorem.

THEOREM 3

Let $x \in \mathbf{R}^N$. If $\|x\|_0 < R = N/(4\Delta(X)^2)$, then x is the unique solution to problem 4.

In other words, LP-minimization can recover any sparse vector x measured by a linear mapping whose kernel has low distortion. To illustrate the condition of this statement, pick A and X as in Theorem 2. We have $\Delta(X) \leq (N \log(N/k)/k)^{1/2}$ and hence $R \geq k/\log(N/k)$.

PROOF: We are given $w = Ax$. Let $S = \text{supp}(x)$. Every solution y to problem 4 satisfies $Ay = w$ and hence is of the form $x + u$ where $u \in \ker(A) = X$. Thus, it suffices to prove if $|S| < R = N/(4\Delta(X)^2)$, then $\|x + u\|_1 > \|x\|_1$ for all $u \in X$.

Indeed,

$$\begin{aligned} \|u + x\|_1 &= \sum_{i \in S} |u_i + x_i| + \sum_{i \notin S} |u_i| && (u_i + x_i = u_i, \text{ if } i \notin S) \\ &\geq \sum_{i \in S} |x_i| - \sum_{i \in S} |u_i| + \sum_{i \in \bar{S}} |u_i| && (\text{Triangle Ineq.}) \\ &= \|x\|_1 + \|u\|_1 - 2 \sum_{i \in S} |u_i|. \end{aligned}$$

On the other hand,

$$\begin{aligned} \sum_{i \in S} |u_i| &\leq \sqrt{|S|} \cdot \|u\|_2 && (\text{Cauchy-Schwartz}) \\ &< \sqrt{R} \cdot \|u\|_2 \\ &\leq \sqrt{R} \cdot \frac{\Delta(X) \|x\|_1}{\sqrt{N}} && (\text{Def. } \Delta) \\ &< \frac{1}{2} \|u\|_2. && (R < N/4\Delta(X)^2) \end{aligned}$$

We conclude that $\|u + x\|_1 > \|x\|_1$. \square

References

- [GLR08] Venkat Guruswami, James R. Lee, and Alexander Razborov. Almost euclidean subspaces of ℓ_1^n via expander codes. In *SODA '08: To appear*, 2008.
- [Ind07] Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1 . In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 615–620, New York, NY, USA, 2007. ACM.