Embeddings of Finite Metric Spaces: The Johnson-Lindenstrauss Flattening Lemma

Daryl Hansen

June 3, 2011

Contents

1 Contents

I am primarily covering a few sections from Matousek's Lectures on Discrete Geometry [1] regarding embeddings of finite metric spaces. In effect, it deals with the manipulation of data sets into forms that imply structure within the data.

I first discuss the possibility of different types of such processes, and go on to discuss a particularly useful result, the Johnson-Lindenstrauss Flattening Lemma. If some constant error coefficient is allowed, this lemma allows for the representation of arbitrary relationships in $O(log(n))$ space.

2 Metric Spaces

2.1 Definition

A metric space is defined as a pair, $X = (A, \rho)$, where:

- A: a set
- ρ: a *metric* mapping $(A \times A) \to [0, \infty)$

and where ρ satisfies certain properties:

- $\rho(x, y) = 0$ iff $x = y$
- Symmetry: $\rho(x, y) = \rho(y, x)$
- Triangle inequality: $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$

Notice that if X is a finite set of n points, it is possible to represent this metric space as $\binom{n}{2}$ numbers representing each $\rho(x_i, x_j)$ value, where $i < j$. However, there are also potentially more useful, and often quite familiar, infinite metric spaces.

2.2 p-Norms

The p-norm of a point $x \in \Re^n$, as seen in previous quarters, is defined as

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|)^{1/p}$$

where $p \geq 1$.

In the case that $p = 2$, for example, this becomes the Euclidean norm.

The pair $(\Re^d, \|x\|_p)$ is denoted as ℓ_p^d. Thus, with this notation Euclidean 3-space is ℓ_2^3.

2.3 Hilbert Spaces

In addition to finite-dimensional spaces, there are times where spaces with unrestricted dimensions are valuable.

Define ℓ_p to be the p-norm metric attached to the set of infinite sequences x where $x = \{x_i\}_{i=1}^{\infty}$. When the 2-norm is used, this infinite-dimensional space ℓ_2 is called the Hilbert space.

3 Embeddings

In this section, consider two metric spaces, $X = (A, \rho)$ and $Y = (B, \sigma)$.

An embedding is a function that maps one metric space into another; that is to say, an f such that $f : X \to Y$.

There are two particular types of embeddings of interest in this paper: isometric embeddings, and the more general D-embeddings.

3.1 Isometric Embeddings

Isometric embeddings have the property that the distance between points are equal to the distance between their images. More formally, if $f : X \to Y$, then f is isometric iff for all $x, y \in A$,

$$\sigma(f(x), f(y)) = \rho(x, y)$$

However, these embeddings do not always exist. For example, a tetrahedron cannot be embedded into a plane with the distance between each vertex preserved. However, it might still be possible to embed it while preserving distances within some fixed error bound D. This D is called the *distortion* of f.

3.2 Distortion and the Lipschitz Norm

If $f : X \to Y$, then the *expansion* of f is the maximal amount by which f increases the distance between two points.

Similarly, the *contraction* of f is the largest ratio by which f decreases the distance between two points.

Define the *distortion* of f to be the *expansion*$\times$*contraction* of f. As a side note, notice that this has the property that any constant scaling of f has no effect on its distortion.

These definitions can be seen a bit more precisely with the use of the Lipschitz norm. Define

$$\|f\|_{Lip} = sup\{\sigma(f(x), f(y))/\rho(x, y) : x, y \in A, x \neq y\}$$

The expansion of f can thus be expressed as $\|f\|_{Lip}$, and the contraction of f is the expansion of its inverse, $\|f^{-1}\|_{Lip}$, while the contraction is now $\|f\|_{Lip} \cdot \|f^{-1}\|_{Lip}$.

3.3 D-Embeddings

A D-embedding can now be defined as an embedding f that has at most a distortion of D.

Through a bit of algebraic manipulation, this is equivalent to there existing some r that satisfies the following equation for each (x, y) pair in A:

$$r \cdot \rho(x, y) \leq \sigma(f(x), f(y)) \leq D \cdot r \cdot \rho(x, y)$$

With this terminology under our belts, it is now possible to go on to the main theorem: that it is possible to embed an $(n + 1)$-point set within ℓ_2^n into ℓ_2^k with only constant distortion, where $k = O(\log(n))$.

First, we examine some useful theorems that will help prove the flattening lemma.

4 Measure concentration

Let S^{n-1} be the unit sphere in $\Re^n$. (As a side note, notice that S^{n-1} is an $(n - 1)$-dimensional surface of the n-dimensional unit ball).

Then, define $P[A]$ on a set $A \subseteq S^{n-1}$ to be the measure of A, normalized such that $P[S^{n-1}] = 1$. This P can be thought of as either the *proportion* of S^{n-1} in A, or equivalently as the *probability measure* of A.

The motivation for this definition is that if we pick a random vector $x \in S^{n-1}$ with uniform distribution, $Prob[x \in A]$ is equivalent to $P[A]$.

In high dimensions, points on the sphere become heavily concentrated. That means that if you know the measure of a set, you also have good information about the measure of its neighborhood.

Define the t-neighborhood of a set A to be the set of points within t of A; that is: $A_t = \{y : \|y - x\| \leq t, x \in A\}$.

The foundation of our later estimates is the following theorem:

$$\text{For any } A \subseteq S^{n-1}, \text{ if } P[A] \geq \frac{1}{2}, \text{ then } 1 - P[A_t] \leq 2e^{-t^2 n/2}$$

which is to say that only a quite small proportion of points in S^{n-1} lie far away from these sets. More specifically, we can also apply this inequality to both the upper and lower hemispheres of S^{n-1} to find that in high dimensions, points are heavily centered around the (or actually, any) equator of a sphere.

4.1 Lévy's lemma

Using this theorem, we can now prove Lévy's lemma:

Let the *median* of f be defined as $med(f) = sup\{t \in \Re : P[f \leq t] \leq \frac{1}{2}\}$.

Then if we have a function $f : s^{n-1} \rightarrow \Re$ that has a Lipschitz norm of 1 (that is, $\|f(x) - f(y)\| \leq x - y$), Lévy's lemma tells us that:

$$P[f > med(f) + t] \leq 2e^{-t^2 n/2} \text{ and } P[f < med(f) + t] \leq 2e^{-t^2 n/2}$$

This can be easily proved using the previous theorem, as $A = \{x \in S^{n-1} : f(x) \leq med(f)\}$ has the property that $P[A] \geq \frac{1}{2}$.

4.2 Concentration of the length of the projection

For a fixed unit vector $x \in S^{n-1}$, let

$$f(x) = \sqrt{x_1^2 + x_2^2 + \ldots + x_k^2}$$

be the length of the projection of x onto the subspace L_0 spanned by the first k dimensions. For a random choice of $x \in S^{n-1}$, this $f(x)$ is is sharply concentrated around a suitable number $m = m(n, k)$:

$$P[|f(x) - m| \geq t] \leq 4e^{-t^2 n/2}$$

for any fixed t.

The distance between projections is never larger than the distance between the original points, so this f is 1-Lipschitz. If we choose $m = med(f)$, Lévy's lemma is then applicable to directly give us the estimate above.

We can also find an approximation for m, namely that if $k \geq 10 \ln(n)$, then $m \geq \frac{1}{2}\sqrt{k/n}$.

As a side note, this lemma also holds when x is held fixed and L is taken as a random k-dimensional subspace of ℓ_2^n.

4.3 Proof

This lower bound can be found by comparing m^2 and $E[f^2]$. For any x, $\|x\|^2 = 1$, so since each coordinate x_i is dealt with symmetrically, $E[x_i^2] = \frac{1}{n}$. This means that $E[f^2] = \frac{k}{n}$.

$E[f^2]$ is the integral of $f(x) * Prob[x]$ over all $x \in S^{n-1}$, and this integral can be split into two parts: $A = \{x : f(x) \leq m + t\}$ and $B = \{x : f(x) >$

$m+t\}$. Approximating the integral over A by $P[A]*\max_{x\in A}(f^2(x))$, and B by $P[B]*\max_{x\in B}(f^2(x))$, we find that

$$\frac{k}{n} \leq E[f^2] \leq P[A]*(m+t)^2+P[B]*\max_{x\in B}(f^2(x)) \leq (m+t)^2+P[B] \leq (m+t)^2+2e^{-t^2n/2}$$

By using the fact that $k \geq 10ln(n)$, and taking $t = \sqrt{k/5n}$, this above inequality can be manipulated to compute that $m \geq \frac{1}{2}\sqrt{k/n}$.

5 Johnson-Lindenstrauss Flattening Lemma

The flattening lemma is as follows:

Let X be an n-point space in ℓ_2, and fix $\epsilon \in [0,1]$. Then there exists a $(1+\epsilon)$-embedding of X into ℓ_2^k, where $k = O(\epsilon^{-2}\log(n))$.

5.1 Proof of Flattening Lemma

Assume n is large. Set $k = 200\epsilon^{-2}\ln(n)$, and assume $k < n$. (Otherwise, there is no embedding necessary).

Take L to be a random k-dimensional subspace of ℓ_2^n, obtained by rotating L_0, and let f be the orthogonal projection from X onto L.

Fix $x, y \in X$. Apply the projection lemma above to:

$$|(\|f(x-y)\|) - m| \leq \frac{\epsilon}{3}m$$

to produce an upper bound on the probability that this equation is violated of:

$$4e^{-\epsilon^2m^2n/18} \leq 4e^{-\epsilon^2k/72} \leq n^{-2}$$

The counting sieve states that the probability of a union of events is at most the sum of those probabilities. Thus, since there are fewer than n^2 pairs of points $x, y \in X$, the probability that for a given L there is some pair that violates the equation above is $\leq n^2 \cdot n^{-2} = 1$, so there must be some L_i in the probability space for which every pair x, y satisfies $|(\|f(x-y)\|) - m| \leq \frac{\epsilon}{3}m$.

For this L_i,

$$(1+\frac{\epsilon}{3})m \leq \|f(x-y)\| \leq (1-\frac{\epsilon}{3})m$$

$$r \cdot \|x-y\| \cdot m \leq \|f(x-y)\| \leq r \cdot \|x-y\| \cdot \frac{1+\epsilon/3}{1-\epsilon/3} \cdot m$$

where r is a suitable constant. Since f is linear, $f(x-y) = f(x) - f(y)$, and since $\epsilon \leq 1$, $\frac{1+\epsilon/3}{1-\epsilon/3} < 1+\epsilon$.

Thus,

$$r \cdot \|x-y\| \cdot m \leq \|f(x) - f(y)\| \leq r \cdot \|x-y\| \cdot (1+\epsilon) \cdot m$$

and f is a $(1+\epsilon)$-embedding onto this L_i, proving the Johnson-Lindenstrauss Flattening Lemma.

6 Applications

There are many fields where computations are done with high-dimensional data sets that are being used to simulate complex processes, such as biology, physics, astronomy, and meteorology. With efficient techniques to find the k-embeddings ensured by the flattening lemma, these systems can be improved in both data storage and time complexity when dealing with problems that do not require exact calculations. Quick matrix multiplication, for example, is quite useful, and can be approximated very quickly by instead multiplying the embeddings of these matrices.

References

[1] Jiří Matoušek, *Lectures in Discrete Geometry.* Springer-Verlag New York, Inc., 2002.