Johnson Lindenstrauss for a k-dimensional subspace

Dylan Satow

10/30/23

Lemma Given a linear map $f : V \to W$ that approximately preserves the magnitudes of and distances between two orthogonal unit vectors $u, v \in V$ with some tolerance $0 < \epsilon < 1$, we can say that

$$-2\epsilon \leq \langle f(u), f(v) \rangle \leq 2\epsilon$$

Proof. First, due to the fact that f approximately preserves $||u - v||$ by definition, we can show one side of the desired inequality using

$$\begin{aligned}
||f(u) - f(v)||^2 &\geq (1 - \epsilon)||u - v||^2 \\
\langle f(u) - f(v), f(u) - f(v) \rangle &\geq (1 - \epsilon)\langle u - v, u - v \rangle \\
||f(u)||^2 + ||f(v)||^2 - 2\langle f(u), f(v) \rangle &\geq (1 - \epsilon)(||u||^2 + ||v||^2 - 2\langle u, v \rangle) \\
-2\langle f(u), f(v) \rangle &\geq (1 - \epsilon)(||u||^2 + ||v||^2 - 2\langle u, v \rangle) - ||f(u)||^2 - ||f(v)||^2 \\
\langle f(u), f(v) \rangle &\leq \frac{1}{2}((\epsilon - 1)(||u||^2 + ||v||^2 - 2\langle u, v \rangle) + ||f(u)||^2 + ||f(v)||^2)
\end{aligned}$$

Now, we can use the fact that

$$||f(u)||^2 + ||f(v)||^2 \leq (1 + \epsilon)||u||^2 + (1 + \epsilon)||v||^2$$

we can write

$$\begin{aligned}
\langle f(u), f(v) \rangle &\leq \frac{1}{2}(\epsilon||u||^2 + \epsilon||v||^2 - ||u||^2 - ||v||^2 - 2(\epsilon - 1)\langle u, v \rangle + (1 + \epsilon)(||u||^2 + ||v||^2)) \\
&\leq \frac{1}{2}(\epsilon||u||^2 + \epsilon||v||^2 - ||u||^2 - ||v||^2 - 2(\epsilon - 1)\langle u, v \rangle + ||u||^2 + ||v||^2 + \epsilon||u||^2 + \epsilon||v||^2) \\
&\leq \frac{1}{2}(2\epsilon(||u||^2 + ||v||^2) + 2(1 - \epsilon)\langle u, v \rangle)
\end{aligned}$$

Because v and u are orthogonal unit vectors, we can say $\langle u, v \rangle = 0$ and $||u|| = ||v|| = 1$ giving us one side of the result

$$\langle f(u), f(v) \rangle \leq 2\epsilon$$

The other side can be shown similarly.

$$\begin{aligned}
||f(u)-f(v)||^2 &\le (1+\epsilon)||u-v||^2 \\
||f(u)||^2+||f(v)||^2-2\langle f(u),f(v)\rangle &\le (1+\epsilon)(||u||^2+||v||^2-2\langle u,v\rangle) \\
-2\langle f(u),f(v)\rangle &\le (1+\epsilon)(||u||^2+||v||^2-2\langle u,v\rangle)-||f(u)||^2-||f(v)||^2 \\
\langle f(u),f(v)\rangle &\ge \frac{1}{2}((-\epsilon-1)(||u||^2+||v||^2-2\langle u,v\rangle)+||f(u)||^2+||f(v)||^2) \\
\langle f(u),f(v)\rangle &\ge \frac{1}{2}((-\epsilon-1)(||u||^2+||v||^2-2\langle u,v\rangle)+(1-\epsilon)(||u||^2+||v||^2)) \\
\langle f(u),f(v)\rangle &\ge \frac{1}{2}(-2\epsilon(||u||^2+||v||^2)+2(1-\epsilon)\langle u,v\rangle) \\
\langle f(u),f(v)\rangle &\ge -2\epsilon
\end{aligned}$$

Giving us our result

$$-2\epsilon \le \langle f(u),f(v)\rangle \le 2\epsilon$$

■

Theorem If we have some n-dimensional linear subspace S of $\mathbb{R}^d$ with orthonormal basis $v_1,\ldots,v_n$, then for any $0<\epsilon<1$ and some $k\in\mathbb{Z}^+$ where $4(\epsilon^2/2-\epsilon^3/3)^{-1}\ln(n+1)\le k<d$, there exists some linear map $f:\mathbb{R}^d\to\mathbb{R}^k$ such that for any $u,w\in S$ where $u=a_1v_1+\cdots+a_nv_n$ and $w=b_1v_1+\cdots+b_nv_n$ we can write

$$(1-\epsilon)||u-w||^2-2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(a_i-b_i)(a_j-b_j)\le ||f(u)-f(w)||^2\le (1+\epsilon)||u-w||^2+2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(a_i-b_i)(a_j-b_j)$$

Furthermore, this map can be found in random polynomial time.

Proof. First, choose n linearly independent unit vectors that act as an orthonormal basis of S

$$v_1,\ldots,v_n\in S$$

Using the Johnson-Lindenstraus Theorem with some $0<\epsilon<1$ and $n+1$ points $v_1,\ldots,v_n,0$ we can say there exists some linear map $f:\mathbb{R}^d\to\mathbb{R}^k$ such that the distances between the vectors $v_1,\ldots,v_n$ are approximately preserved for some integer $k\ge 4(\epsilon^2/2-\epsilon^3/3)^{-1}\ln(n+1)$. Clearly, the magnitudes of each $v_1,\ldots,v_n$ must be approximately preserved due to the inclusion of 0.

Now, for any two vectors $u,w\in S$, their difference $u-w$ must clearly belong to S as well. Thus, if we can show that the magnitudes of all elements $u\in S$ are preserved, then the interpoint distances for all points in S must also be preserved.

First, we can say $u = a_1v_1 + \cdots + a_nv_n$ because $v_1, \ldots, v_n$ spans S. Now, using the linearity of f,we can write

$$\begin{aligned}
||f(u)||^2 &= ||f(a_1v_1 + \cdots + f(a_nv_n))||^2 \\
&= ||a_1f(v_1) + \cdots + a_nf(v_n)||^2 \\
&= \langle a_1f(v_1) + \cdots + a_nf(v_n), a_1f(v_1) + \cdots + a_nf(v_n)\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\langle a_if(v_i), a_jf(v_j)\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\langle f(v_i), f(v_j)\rangle \\
&= a_1^2||f(v_1)||^2 + \cdots + a_n^2||f(v_n)||^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_ia_j\langle f(v_i), f(v_j)\rangle
\end{aligned}$$

Now, using the lemma, the bound for f, and the fact that $v_1, \ldots, v_n$ are unit, we can write

$$(1-\epsilon)(a_1^2 + \cdots + a_n^2) - 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_ia_j \leq ||f(u)||^2 \leq (1+\epsilon)(a_1^2 + \cdots + a_n^2) + 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_ia_j$$

$$(1-\epsilon)||u||^2 - 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_ia_j \leq ||f(u)||^2 \leq (1+\epsilon)||u||^2 + 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_ia_j$$

We have succesfully created a bound for any vector $u \in S$. Now, because $u, w \in S \rightarrow u - w \in S$, if we define some $u = a_1v_1 + \cdots a_nv_n$ and $v = b_1v_1 + \cdots + b_nv_n$, where $v_1, \ldots, v_n$ is an orthonormal basis of S, we can write

$$(1-\epsilon)||u-w||^2 - 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(a_i-b_i)(a_j-b_j) \leq ||f(u)-f(w)||^2 \leq (1+\epsilon)||u-w||^2 + 2\epsilon\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(a_i-b_i)(a_j-b_j)$$

Giving us our result and finishing the proof.

∎