The Johnson-Lindenstrauss Lemma

Kevin (Min Seong) Park

MAT477

Introduction

The Johnson-Lindenstrauss Lemma was first introduced in the paper ”Extensions of Lipschitz mappings into a Hilbert Space” by William B. Johnson and Joram Lindenstrauss published 1984 in Contemporary Mathematics. The Theorem is as follows.

1. *Johnson-Lindenstrauss Lemma*
Fix $0 < \epsilon < 1$, *let* $V = \{x_i : i = 1, ...M\} \subset \mathbb{R}^m$ *be a set of points in* $\mathbb{R}^m$ *If* $n \geq \frac{c}{\epsilon^2} \log M$ *then there exists a linear map* $A : \mathbb{R}^m \to \mathbb{R}^n$ *such that for all* $i \neq j$

$$1 - \epsilon \leq \frac{\|A(x_i) - A(x_j)\|}{\|x_i - x_j\|} \leq 1 + \epsilon.$$

The Theorem states that after fixing an error level, one can map a collection of points from one Euclidean space (no matter how high it's dimension m is) to a smaller Euclidean space while only changing the distance between any two points by a factor of $1 \pm \epsilon$. The dimension of the image space is only dependent on the error and the number of points. Given that the dimension is very large, one can achieve significant dimension reduction, which has applications in data analysis and computer science.

There are two proofs that use applications of Gaussian distributions. One will use the Gaussian concentration inequality for Lipschitz functions, and the other will use the comparison inequalities called the Gaussian Min-Max Theorems.

Both proofs have the same probabilistic approach: Prove that the probability of a linear map satisfying the conditions of the theorem is positive.

This is done in the following way: Construct a random matrix from $\mathbb{R}^m$ to $\mathbb{R}^n$ which is an m by n matrix where the entries g_{ij} are independent standard Gaussian variables.

$$G = \begin{bmatrix} g_{11} & \cdots g_{1m} \\ \vdots & \vdots \\ g_{n1} & \cdots g_{nm} \end{bmatrix} \text{ where } g_{ij} \text{ are i.i.d standard Gaussian variables.}$$

Then the random matrix maps the difference of two points, and the image will be Random Vector which follows a Gaussian distribution. From there, the two proofs branch off:

The first will use the Gaussian concentration inequality on Lipschitz functions to provide an lower bound on the probability that the distance between any two points changes by only a factor of $1 \pm \epsilon$. The lower bound will be dependent on n, ϵ, and M. The dependence on M will come from having to compare all possible pair of points. Thus by choosing n nicely, the lower bound will be strictly greater than 0 proving that the probability of there existing such a map satisfying the lemma is positive so there must exist such a map.

The second proof will calculate the expectation that the maximum change in distance is less than $1 + \epsilon$ and the minimum change in distance is greater than $1 - \epsilon$. Then, by defining two random bilinear forms and comparing their expectations with the min-max theorems, the result follows.

1 Proof by Gaussian Concentration

We have our random Gaussian matrix G. Denote $y = Gx$. y is the image of the linear map G at a point $x \in V$. Computing y explicitly,

$$\begin{aligned} y &= \begin{bmatrix} g_{11} & \cdots g_{1m} \\ \vdots & \vdots \\ g_{m1} & \cdots g_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \\ &= \begin{bmatrix} g_{11}x_1 + & \cdots + g_{1m}x_m \\ \vdots & \vdots \\ g_{n1}x_1 + & \cdots + g_{nm}x_m \end{bmatrix}. \end{aligned}$$

If we apply the random linear map G to two points $x^p, x^q \in V$, then to prove the Johnson Lindenstrauss Lemma, we want

$$\mathbb{P}\left(\forall x^p, x^q \in V \ : \ (1-\epsilon)\|x^p - x^q\| \le \|y^p - y^q\| \le (1+\epsilon)\|x^p - x^q\|\right) > 0. \tag{1}$$

This shows that there must exist a linear map that satisfies the Johnson Lindenstrauss Lemma.

To set up (1), we investigate y. By the stability property of the sum of i.i.d Gaussian variables, y is a Gaussian random vector where the y_i are independent Gaussian random variables with mean 0 and variance $\sum_{i=1}^m x_i^2$.

$$y =_d \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} \text{ where } g_1, \dots g_n \text{ are i.i.d } \sim N(0, \sum_{i=1}^m x_i^2).$$

Thus we can say that

$$y =_d \|x\| z \ , \ z = (z_1, \dots, z_n) \ , \ z_j \text{ are i.i.d } \sim N(0,1).$$

Now we repeat the above calculation to two different vectors x^p and x^q in V and obtain y^p and y^q. Their difference has a distribution expressed as

$$y^p - y^q =_d \|x^p - x^q\| z$$

Taking Euclidean norm on both sides, we have an expression for the distance between two points in the lower dimensional space:

$$\|y^p - y^q\| =_d \|x^p - x^q\| \|z\|. \tag{2}$$

We want to calculate the probability that this distance is within a factor of $1 \pm \epsilon$ of the original distance. That is, we want to calculate the expression

$$\mathbb{P}((1-\epsilon)\|x^p - x^q\| \le \|y^p - y^q\| \le (1+\epsilon)\|x^p - x^q\|). \tag{3}$$

To do so, we use the Gaussian concentration bound for Lipschitz functions.

Lemma 1. *Consider a Lipschitz function $F : \mathbb{R}^m \to \mathbb{R}$ such that for some $L > 0$,*

$$\|F(x) - F(y)\| \le L\|x - y\| \text{ for all } x, y \in \mathbb{R}^m$$

Let $g = (g_i)_{i \le m}$ be a standard Gaussian vector in $\mathbb{R}^m$. Then for any $t \ge 0$,

$$\mathbb{P}\Big(\|F(g) - \mathbb{E}F(g)\| \geq t\Big) \leq 2exp\left(-\frac{t^2}{4L^2}\right)$$

For the proof, refer to Section 7 of Gaussian Distributions with Applications. [1]

To apply the lemma, we need a Lipschitz function. The next two lemmas show us that the norm function is a Lipschitz function.

Lemma 2.

$$\|v\| = \sup\Big\{\sum_i^m \alpha_i v_i \,:\, \|\alpha\| = 1\Big\} \text{ for } v \in \mathbb{R}^m$$

Proof. The $\geq$ comes from the Cauchy-Schwarz Inequality. For any α with $\|\alpha\| = 1$

$$\|\langle \alpha, v\rangle\| \leq \|\alpha\|\|v\|$$

For the $\leq$, take $\alpha = \frac{v}{\|v\|}$. Then $\langle \alpha, v\rangle = \|v\|$ □

Lemma 3. *Let A be a bounded set in $\mathbb{R}^m$. Consider the function $F : \mathbb{R}^m \to \mathbb{R}$ defined by*

$$F(x) = \sup_{a\in A}\langle a, x\rangle.$$

Then F is a Lipschitz Function.

Proof.

$$\begin{aligned}
\|F(x) - F(y)\| &= \|\sup_{a\in A}\langle a, x\rangle - \sup_{a\in A}\langle a, y\rangle\| \\
&= \|\sup_{a\in A}(a_1x_1 + \cdots + a_mx_m) - \sup_{a\in A}(a_1y_1 + \cdots + a_my_m)\| \\
&\leq \|\sup_{a\in A}(a_1(x_1 - y_1) + \cdots + a_m(x_m - y_m))\| \\
&= \sup_{a\in A}\|(a_1(x_1 - y_1) + \cdots + a_m(x_m - y_m))\| \\
&\leq \sup_{a\in A}\|a\|\|x - y\|
\end{aligned}$$

□

So from (2) we express $\|z\|$ as

$$F(z) = \|z\| = \sup_{\|\alpha\|=1}\Big\{\sum_i^m \alpha_i z_i\Big\}$$

which is a Lipschitz function. The bounded set we take the supremum over is $A = \{a \in \mathbb{R}^m \ : \ \|a\| = 1\}$ and so the Lipschitz bound we use is $L = 1$.

By Gaussian Concentration

$$\mathbb{P}\left(\ \big| \ \|z\| - \mathbb{E}\|z\| \ \big| \geq t \right) \leq 2e^{-t^2/4}. \tag{4}$$

We manipulate (4) to obtain a lower bound for (3). We take the complement to switch the inequalities in (4).

$$\mathbb{P}\left(\ \big| \ \|z\| - \mathbb{E}\|z\| \ \big| \leq t \right) \leq 1 - 2e^{-t^2/4}.$$

We notice that $\mathbb{E}\|z\|$ is a constant term so we can perform a change of variables $t = \epsilon\mathbb{E}\|z\|$.

$$\mathbb{P}\left(1 - \epsilon \leq \frac{\|z\|}{\mathbb{E}\|z\|} \leq 1 + \epsilon \right) \leq 1 - 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

By multiplying the equation inside the probability by $\|x^p - x^q\|$

$$\mathbb{P}\left((1-\epsilon)\|x^p - x^q\| \leq \frac{\|y^p - y^q\|}{\mathbb{E}\|z\|} \leq (1+\epsilon)\|x^p - x^q\| \right) \geq 1 - 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}. \tag{5}$$

However this is not exactly the probability we want to calculate. We want to compare $\|x^p - x^q\|$ with $\|y^p - y^q\|$, not $\frac{\|y^p - y^q\|}{\mathbb{E}\|z\|}$. In order to get rid of the denominator term $\mathbb{E}\|z\|$, we alter the linear map G by dividing it by the constant $\mathbb{E}\|z\|$. Formally, define $\hat{G}$ as

$$\hat{G} = \frac{1}{\mathbb{E}\|z\|} G.$$

Then if we let $\hat{y} = \hat{G}x$,

$$\hat{y}^p - \hat{y}^q =_d \frac{\|x^p - x^q\|}{\mathbb{E}\|z\|} z.$$

Repeating all the calculations that led to (5), we have that

$$\mathbb{P}((1-\epsilon)\|x^p - x^q\| \leq \|\hat{y}^p - \hat{y}^q\| \leq (1+\epsilon)\|x^p - x^q\|) \geq 1 - 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}$$

Let us pretend that our original linear map G took into account the denominator $\mathbb{E}\|z\|$ to obtain the lower bound for (3)

$$\mathbb{P}\left((1-\epsilon)\|x^p - x^q\| \leq \|y^p - y^q\| \leq (1+\epsilon)\|x^p - x^q\|\right) \geq 1 - 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}. \quad (6)$$

To go from (6) to (1), we have to consider the probability for any two points in V. To do so, we consider the complement of the event and use the union bound. We reformulate (6) into set notation by letting A_{pq} be the event that

$$(1-\epsilon)\|x^p - x^q\| \leq \|y^p - y^q\| \leq (1+\epsilon)\|x^p - x^q\| \text{ occurs.}$$

To calculate the probability that every pair of points satisfies the inequality, we are calculating the probability of the intersection of all the A_{pq}'s. Explicitly we are calculating $\mathbb{P}(\cap A_{pq})$ where the intersection is over all possible choice of pairs in V. To do so, first notice that

$$\mathbb{P}(A_{pq}) \geq 1 - 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

Then by taking the complement,

$$\mathbb{P}(A_{pq}^{\complement}) \leq 2e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

Notice that there are M points in V so the number of pairs is $\binom{M}{2} \leq \frac{M^2}{2}$. Thus by the union bound $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$, we have that

$$\mathbb{P}(\cup A_{pq}^{\complement}) \leq 2\binom{M}{2} e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}} \leq M^2 e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

By DeMorgan's law, $\cup A_{pq}^{\complement} = (\cap A_{pq})^{\complement}$

$$\mathbb{P}((\cap A_{pq})^{\complement}) \leq M^2 e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

Take complements again to obtain

$$\mathbb{P}(\cap A_{pq}) \geq 1 - M^2 e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}.$$

This is exactly a lower bound for (1):

$$\mathbb{P}(\forall x^p, x^q \in V \ : (1-\epsilon)\|x^p - x^q\| \leq \|y^p - y^q\| \leq (1+\epsilon)\|x^p - x^q\|) \geq 1 - M^2 e^{-\frac{\epsilon^2(\mathbb{E}\|z\|)^2}{4}}$$

Finally, to satisfy (1) and complete the proof, we need the condition that $1 - M^2 e^{-\frac{\epsilon^2 (\mathbb{E}\|z\|)^2}{4}} > 0$. This condition will lead us to the minimum dimension of the codomain. Calculating

$$
\begin{aligned}
1 - M^2 e^{-\frac{\epsilon^2 (\mathbb{E}\|z\|)^2}{4}} &> 0. \\
\log(\frac{1}{M^2}) &> -\frac{\epsilon^2}{4}(\mathbb{E}\|z\|)^2. \\
(\mathbb{E}\|z\|)^2 &> \frac{8}{\epsilon^2}\log M.
\end{aligned}
$$

The condition we need to satisfy is $(\mathbb{E}\|z\|)^2 > \frac{8}{\epsilon^2}\log M$. We use the fact that $\mathbb{E}\|z\| \geq k\sqrt{n}$ where k is a constant and n is the dimension of the random vector z. So if $n \geq \frac{8c}{\epsilon^2}\log M$, where c is a constant, then the condition $1 - M^2 e^{-\frac{\epsilon^2 (\mathbb{E}\|z\|)^2}{4}} > 0$ is satisfied. Therefore, if $n > \frac{8c}{\epsilon^2}\log M$, we have that

$$
\mathbb{P}\left(\forall x^p, x^q \in V \;:\; (1-\epsilon)\|x^p - x^q\| \leq \|y^p - y^q\| \leq (1+\epsilon)\|x^p - x^q\|\right) > 0
$$

so there must exist a linear map G_0 that satisfies

$$
\forall x^p, x^q \in V \;:\; (1-\epsilon)\|x^p - x^q\| \leq \|G_0(x^p) - G_0(x^q)\| \leq (1+\epsilon)\|x^p - x^q\|.
$$

This completes the proof.

2 Gaussian Min-Max Comparison

The Gaussian Min-Max Comparison Inequalities are the Slepian's and Gordon's Inequalities.

Lemma 4. *Slepian's Inequality Let $X(i)$ and $Y(i)$ be two Gaussian vectors such that*

$$
\mathbb{E}(Y(i) - Y(i'))^2 \geq \mathbb{E}(X(i) - X(i'))^2
$$

Then

$$
\mathbb{E}\max_i Y(i) \geq \mathbb{E}\max_i X(i)
$$

Lemma 5. *Gordon's Ineqaulity Let* $(X(i,j))$ *and* $(Y(i,j)$ *be two Gaussian vectors such that*

$$\mathbb{E}(Y(i,j)-Y(i,j'))^2 \leq \mathbb{E}(X(i,j)-X(i,j'))^2$$
$$\mathbb{E}(Y(i,j)-Y(i',j'))^2 \geq \mathbb{E}(X(i,j)-X(i',j'))^2$$

Then

$$\mathbb{E}\min_i \max_j Y(i,j) \leq \mathbb{E}\min_i \max_j X(i,j)$$

The proofs can be found in Chapter 9 of [1], but here is a short sketch of both proofs. First, give a smooth approximation of the max function (or min max function). Then make an interpolation between X and Y and show that the function is increasing (or decreasing) by differentiating using Gaussian integration by parts.

We will use these two lemmas to prove the Johnson Lindenstrauss Lemma. First let us set up what we would like to prove and where the inequalities come to prove the Johnson Lindenstrauss Lemma. Define T as

$$T = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|} : i \neq j \right\} \text{ ; where } x_i, x_j \in V.$$

T is a subset of less than M^2 points on the n-dimensional sphere. The advantage of using T will be that to prove the Johnson- Lindenstrauss lemma, all we need to show is that there is a linear map G_0 such that

$$1-\epsilon \leq \|G_0(t)\| \leq 1+\epsilon \quad \forall t \in T.$$

We can express this as

$$\max_{t\in T} \|G_0(t)\| \leq 1+\epsilon \quad \min_{t\in T} \|G_0(t)\| \geq 1-\epsilon. \tag{7}$$

Of course the minimum is less than the maximum, hence we rewrite (7) as

$$1-\epsilon \leq \min_{t\in T} \|G_0(t)\| \leq \max_{t\in T} \|G_0(t)\| \leq 1+\epsilon$$

As long as $\min_{t\in T} \|G_0(t)\| \leq 1$ (which is not a problem as we can always scale our matrix by a constant), we can further rewrite (7) as

$$\frac{\max_{t\in T}\|G_0(t)\|}{\min_{t\in T}\|G_0(t)\|} \leq \frac{1+\epsilon}{1-\epsilon}$$

and even further as

$$\max_{t\in T}\|G_0(t)\| - \frac{(1+\epsilon)}{(1-\epsilon)}\min_{t\in T}\|G_0(t)\| \leq 0. \tag{8}$$

So if we show (8) then we are finished. Notice that if we are under our random matrix set up G from before, then we can see

$$\max_{t\in T}\|G(t)\| - \frac{(1+\epsilon)}{(1-\epsilon)}\min_{t\in T}\|G(t)\|$$

as a random variable. The probability space where this random variable is defined on are n by m matrices. We can consider the expectation of this random variable

$$\mathbb{E}\max_{t\in T}\|G(t)\| - \frac{(1+\epsilon)}{(1-\epsilon)}\mathbb{E}\min_{t\in T}\|G(t)\|$$

and if the expectation is less than 0, then there must be a matrix G_0 which satisfies (8). Thus, we are left to show that

$$1-\epsilon \leq \mathbb{E}\min_{t\in T}\|G(t)\| \leq \mathbb{E}\max_{t\in T}\|G(t)\| \leq 1+\epsilon.$$

To prove the lemma, we will provide bounds for $\mathbb{E}\min_{t\in T}\|G(t)\|$ below and $\mathbb{E}\max_{t\in T}\|G(t)\|$ above by using the Gaussian min-max inequalities. We will compare $\|G(t)\|$ to a Gaussian vector for which the bounds on its minimum and maximum are easy to compute. Then we give conditions for which the bounds will be satisfied. Let us get started.

We use the same linear map G as in the previous proof. We repeat it here:

$$G = \begin{bmatrix} g_{11} & \cdots g_{1m} \\ \vdots & \vdots \\ g_{n1} & \cdots g_{nm} \end{bmatrix} \; g_{ij} \text{ are i.i.d } \sim N(0,1).$$

We map any point $t = (t_1, \ldots, t_m) \in T$. The image will be $G(t)$.

$$G(t) = \begin{bmatrix} g_{11} & \cdots g_{1m} \\ \vdots & \vdots \\ g_{n1} & \cdots g_{nm} \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix} = \begin{bmatrix} g_{11}t_1 + & \cdots + g_{1m}t_m \\ \vdots & \vdots \\ g_{n1}t_1 + & \cdots + g_{nm}t_m \end{bmatrix}.$$

Define two Gaussian bilinear forms: Let $t \in T$, $u \in B^m$ the unit ball in $\mathbb{R}^m$, and let $\{h_i\}_{i=1}^n,\ \{g_j\}_{j=1}^m$ all be i.i.d $\sim N(0,1)$

$$X(t,u) = \langle G(t), u\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}t_iu_j \quad Y(t,u) = \sum_{i=1}^{n} t_ih_i + \sum_{j=1}^{m} u_jg_j$$

These can be found in example 3 of Chapter 10 in [1], where $a, b = 1$ and $U = B^m$ The m-dimensional unit ball.

We can motivate the definition of $X(t,u)$. From Lemma 2 in the previous section, if we maximize over $u \in B^m$, then we are calculating $\|G(t)\|$. Then by taking the minimum, over $t \in T$ and taking the expectation, we have that

$$\mathbb{E} \min_{t\in T} \max_{u\in B} X(t,u) = \mathbb{E} \min_{t\in T} \|G(t)\|.$$

Similarly,

$$\mathbb{E} \max_{t\in T} \max_{u\in B} X(t,u) = \mathbb{E} \max_{t\in T} \|G(t)\|.$$

So if we show that the conditions for lemma 4 and 5 are met for $X(t,u)$ and $Y(t,u)$, then we can give bounds for $\mathbb{E}\min_{t\in T} \|G(t)\|$ and $\mathbb{E}\max_{t\in T} \|G(t)\|$ as

$$\mathbb{E} \min_{t\in T} \max_{u\in B} Y(t,u) \le \mathbb{E} \min_{t\in T} \|G(t)\| \le \mathbb{E} \max_{t\in T} \|G(t)\| \le \mathbb{E} \max_{t\in T} \max_{u\in B} Y(t,u). \tag{9}$$

Checking that $X(t,u)$ and $Y(t,u)$ meet the conditions of the Slepian's and Gordon's Inequalities is tedious and will be left until the end. Let us continue with (9).

It is easy to maximize and minimize $Y(t,u)$ over $t \in T$ and $u \in B^m$ because Y is defined as

$$Y(t,u) = \langle t,h\rangle + \langle u,g\rangle$$

so maximizing and minimizing over t and u can be done separately. Explicitly, we can see that the right most term of (9) can be written as

$$\mathbb{E}\max_{t\in T}\max_{u\in B} Y(t,u) = \mathbb{E}\max_{t\in T}\langle t,h\rangle + \mathbb{E}\max_{u\in B}\langle u,g\rangle \tag{10}$$

$$= \mathbb{E}\max_{t\in T}\langle t,h\rangle + \mathbb{E}\|g\| \tag{11}$$

We can also show that the left most term of (8), using a "change of variables" trick, can be written as

$$\mathbb{E}\min_{t\in T}\max_{u\in B} Y(t,u) = \mathbb{E}\|g\| - \mathbb{E}\max_{t\in T}\langle t,h\rangle \tag{12}$$

we will leave the details of (12) to later. From here, we are almost finished. To prove the Johnson Lindenstrauss lemma, all we need to do is to have

$$\mathbb{E}\min_{t\in T}\max_{u\in B} Y(t,u) \geq 1-\epsilon \quad \mathbb{E}\max_{t\in T}\max_{u\in B} Y(t,u) \leq 1+\epsilon$$

and from (11) and (12), all we need is $\mathbb{E}\max_{t\in T}\langle t,h\rangle \leq \epsilon\mathbb{E}\|g\|$ which is satisfied if

$$n \geq \frac{c}{\epsilon^2}\log M.$$

where c is a constant, which is our bound on the dimension in the condition for the Johnson Lindenstrauss lemma. $n \geq \frac{c}{\epsilon^2}\log M$ implying $\mathbb{E}\max_{t\in T}\langle t,h\rangle \leq \epsilon\mathbb{E}\|g\|$ will be left for later. This finishes the proof of the Johnson Lindenstrauss lemma.

Let us fill in all the details:

2.1 Check Conditions for Lemmas 4 and 5

With our Gaussian vectors $X(t,u)$ and $Y(t,u)$, let us show that they satisfy the conditions for Gordon's Inequality (lemma 5). Namely

$$\mathbb{E}(Y(t,u) - Y(t,u'))^2 \leq \mathbb{E}(X(t,y) - X(t,u'))^2$$

and

$$\mathbb{E}(Y(t,u) - Y(t',u'))^2 \geq \mathbb{E}(X(t,u) - X(t',u'))^2.$$

If these conditions are satisfied, then

$$\mathbb{E} \min_{t\in T} \max_{u\in B} Y(t,u) \leq \mathbb{E} \min_{t\in T} \max_{u\in B} X(t,u).$$

For the first condition,

$$\begin{aligned}
\mathbb{E}(X(t,u) - X(t',u'))^2 &= \mathbb{E}(\sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}(t_i u_j - t'_i u'_j))^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}(t_i u_j - t'_i u'_j)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}(t_i u_j)^2 - 2(t_i u_j t'_i u'_j) + (t'_i u'_j)^2 \\
&= \|t\|^2\|u\|^2 + \|t'\|^2\|u'\|^2 - 2\langle t, t'\rangle\langle u, u'\rangle \\
&= \|u\|^2 + \|u'\|^2 - 2\langle t, t'\rangle\langle u, u'\rangle
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(Y(t,u) - Y(t',u'))^2 &= \mathbb{E}(\sum_{i=1}^{n} h_i(t_i - t'_i) + \sum_{j=1}^{m} g_j(u_j - u_j))^2 \\
&= \mathbb{E}(\sum_{i=1}^{n} h_i(t_i - t'_i))^2 + \mathbb{E}(\sum_{j=1}^{m} g_j(u_j - u_j))^2 \\
&= \sum_{i=1}^{n}(t_i - t'_i)^2 + \sum_{j=1}^{m}(u_j - u_j)^2 \\
&= \sum_{i=1}^{n} t_i^2 - 2t_i t'_i + (t'_i)^2 + \sum_{j=1}^{m} u_j^2 - 2u_j u'_j + (u'_j)^2 \\
&= \|t\| + \|t'\| - \langle t, t'\rangle + \|u\| + \|u'\| - 2\langle u, u'\rangle \\
&= 2 - 2\langle t, t'\rangle + \|u\| + \|u'\| - 2\langle u, u'\rangle.
\end{aligned}$$

Subtract these two equations to find that

$$
\begin{aligned}
&\mathbb{E}(Y(t,u) - Y(t',u'))^2 - \mathbb{E}(X(t,u) - X(t',u'))^2 \\
&= 2 - 2\langle t,t'\rangle - 2\langle u,u'\rangle + 2\langle t,t'\rangle\langle u,u'\rangle \\
&= 2(1 - \langle t,t'\rangle)(1 - \langle u,u'\rangle) \\
&\geq 0.
\end{aligned}
$$

The last inequality is a consequence of Cauchy-Schwarz on $\langle t,t'\rangle$ and $\langle u,u'\rangle$ and the fact that t is on the unit sphere, u is in the unit ball.
For the second condition, let $t = t'$. Then we have that $\langle t,t'\rangle = 1$ so $\mathbb{E}(Y(t,u) - Y(t,u'))^2 - \mathbb{E}(X(t,u) - X(t,u'))^2 = 0$. The second condition is satisfied trivially.

Now let us show that $X(t,u)$ and $Y(t,u)$ satisfy the conditions for Slepian's Inequality (lemma 4) for both indexes t and u. All of the work is actually done when showing the conditions for the Gordon's Inequality. If t is fixed, then we want to show

$$\mathbb{E}(Y(t,u) - Y(t,u'))^2 \geq \mathbb{E}(X(t,u) - X(t,u'))^2.$$

We have already shown that equality holds so the condition is satisfied trivially. If u is fixed, then we want to show

$$\mathbb{E}(Y(t,u) - Y(t',u))^2 \geq \mathbb{E}(X(t,u) - X(t',u))^2.$$

We have already shown that

$$
\begin{aligned}
&\mathbb{E}(Y(t,u) - Y(t',u'))^2 - \mathbb{E}(X(t,u) - X(t',u'))^2 \\
&= 2(1 - \langle t,t'\rangle)(1 - \langle u,u'\rangle).
\end{aligned}
$$

If we fix $u = u'$ in the unit ball, we have that $\langle u,u\rangle \leq 1$. Thus

$$
\begin{aligned}
&\mathbb{E}(Y(t,u) - Y(t',u))^2 \geq \mathbb{E}(X(t,u) - X(t',u))^2 \\
&= 2(1 - \langle t,t'\rangle)(1 - \langle u,u'\rangle) \\
&\geq 2(1 - \langle t,t'\rangle) \\
&\geq 0
\end{aligned}
$$

where the last inequality again follows from Cauchy Schwarz.

Then Slepian's Inequality implies that

$$\mathbb{E}\max_{t\in T}\max_{u\in B}X(t,u)\leq\mathbb{E}\max_{t\in T}\max_{u\in B}Y(t,u).$$

Putting the consequences of Gordon's Inequality and Slepian's Inequality together, and the fact that

$$\mathbb{E}\min_{t\in T}\max_{u\in B}X(t,u)\leq\mathbb{E}\max_{t\in T}\max_{u\in B}X(t,u),$$

we have (8):

$$\mathbb{E}\min_{t\in T}\max_{u\in B}Y(t,u)\leq\mathbb{E}\min_{t\in T}\max_{u\in B}X(t,u)\leq\mathbb{E}\max_{t\in T}\max_{u\in B}X(t,u)\leq\mathbb{E}\max_{t\in T}\max_{u\in B}Y(t,u)$$

The first inequality is by Gordon's, and the last inequality is by Slepian's.

2.2 Computing (12)

We want to show that

$$\mathbb{E}\min_{t\in T}\max_{u\in B}Y(t,u)=\mathbb{E}\|g\|-\mathbb{E}\max_{t\in T}\langle t,h\rangle.$$

We know that, when maximizing over u and minimizing over t we can separate the left hand side as

$$\mathbb{E}\min_{t\in T}\max_{u\in B}Y(t,u)=\mathbb{E}\min_{t\in T}\sum_{i=1}^{n}t_ih_i+\mathbb{E}\max_{u\in B}\sum_{j=1}^{m}u_jg_j$$

and $\mathbb{E}\|g\|=\mathbb{E}\max_{u\in B}\sum_{j=1}^{m}u_jg_j$. All that is left to show is

$$\mathbb{E}\min_{t\in T}\sum_{i=1}^{n}t_ih_i=-\mathbb{E}\max_{t\in T}\langle t,h\rangle.$$

We will use a "Change of Variables" where we change h_i to $-h_i$. $-h_i$ is still a Gaussian variable because the Gaussian variables are invariant under rotation. Then

$$\mathbb{E}\min_{t\in T}\sum_{i=1}^{n}t_ih_i=\mathbb{E}\min_{t\in T}\sum_{i=1}^{n}t_i(-h_i)=-\mathbb{E}\max_{t\in T}\sum_{i=1}^{n}t_ih_i=-\mathbb{E}\max_{t\in T}\langle t,h\rangle$$

because the minimum of negative values is the negative of the maximum of the positive values.

2.3 Computing the lower bound on the dimension

We want to show that if $n \geq \frac{c}{\epsilon^2} \log M$ implies that $\mathbb{E} \max_{t \in T} \langle t, h \rangle \leq \epsilon \mathbb{E}\|g\|$, which proves the Johnson Lindenstrauss Lemma.

We use the fact previously mentioned in the first section: $\exists k$ s.t $k\sqrt{n} \leq \mathbb{E}\|g\|$.

We will prove that $\mathbb{E} \max_{t \in T} \langle t, h \rangle \leq 2\sqrt{logM}$. Thus, after some computation,

$$n \geq \frac{k}{\epsilon^2} \log M \Leftrightarrow 2\sqrt{\log M} \leq c\epsilon\sqrt{n}$$

where k and c are constants. So if $n \geq \frac{c}{\epsilon^2} \log M$, then

$$\mathbb{E} \max_{t \in T} \langle t, h \rangle \leq 2\sqrt{logM} \leq c\epsilon\sqrt{n} \leq \epsilon \mathbb{E}\|g\|$$

To show that $\mathbb{E} \max_{t \in T} \langle t, h \rangle \leq 2\sqrt{\log M}$

$$\begin{aligned}
\mathbb{E} \max_{t \in T} \langle t, h \rangle &= \mathbb{E} \max_{t \in T} \sum_{i=1}^{n} t_i h_i \\
&= \mathbb{E} \max_{t \in T} h_t \quad \left[h_t \sim N\left(0, \sum t_i^2\right) = N(0,1) \right] \\
&\leq \mathbb{E} \frac{1}{\beta} \log\left(\sum_{t=1}^{M^2} e^{h_t \beta} \right) \quad \left[\max_i x_i \leq \frac{1}{\beta} \log\left(\sum_{t=1} e^{x_i \beta} \right) \right] \\
&\leq \frac{1}{\beta} \log\left(\sum_{t=1}^{M^2} \mathbb{E} e^{h_t \beta} \right) \quad \text{By Jensen's Inequality} \\
&= \frac{1}{\beta} \log\left(\sum_{t=1}^{M^2} e^{\beta^2/2} \right) \quad \text{because } \mathbb{E} e^g = e^{Var(g)/2} \\
&= \frac{1}{\beta} \log(M^2 e^{\beta^2/2}) \\
&= \frac{1}{\beta} \log M^2 + \frac{\beta}{2}.
\end{aligned}$$

This is true for all β so minimizing the last equation $\beta = \sqrt{2\log M^2}$

$$\mathbb{E}\max_{t\in T}\langle t, h\rangle \leq 2\sqrt{\log M}$$

We have used the fact that $\mathbb{E}e^g = e^{Var(g)/2}$ where g is a Gaussian variable. The calculation is easy and is left as an exercise for the reader.

Conclusion

The original proof as well as the two proofs provided have all been probabilistic. There is a non-zero probability that this map exists. This does not help us in actually constructing such a map. However, over the years, different proofs have been provided where this map can be generated through algorithms in randomized polynomial time, and even further work has been done to derandomize the algorithms.

Appendix A

The final bound on the dimension relied on the fact that $\mathbb{E}\|g\| \geq k\sqrt{n}$ where k is a constant and g is a standard Gaussian vector of dimension n. This fact is not so easy and the details are contained in [2] (Thank you David Miyamoto!). We can give a summary of the calculation here.

$$\mathbb{E}\|g\| = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n}\|x\|e^{-\frac{\|x\|}{2}}dx.$$

By switching to spherical coordinates, using a property of the beta function, and skipping a lot of calculations (all included in [2]), one can show that

$$\mathbb{E}\|g\| = \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}.$$

Then, one shows by induction and properties of the Gamma function that for all n,

$$\frac{n}{\sqrt{n+1}} \leq \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \leq \sqrt{n}$$

and finally, one can verify that there exists a constant k such that for all n,

$$k\sqrt{n} \leq \frac{n}{\sqrt{n+1}}.$$

Appendix B

There is a variant of the Johnson Lindenstrauss Lemma proven in [4] and [5]. The difference is that they prove for all $i \neq j$

$$(1-\epsilon)\|x_i - x_j\|^2 \leq \|A(x_i) - A(x_j)\|^2 \leq (1+\epsilon)\|x_i - x_j\|^2 \tag{13}$$

where $x_i, x_j \in V$. This is already a better linear map than the one shown to exist in the original version of the Johnson Lindenstrauss because. By taking square roots, the above shows that

$$\sqrt{1-\epsilon}\|x_i - x_j\| \leq \|A(x_i) - A(x_j)\| \leq \sqrt{1+\epsilon}\|x_i - x_j\|$$

and $1 - \epsilon < \sqrt{1-\epsilon}$ and $\sqrt{1+\epsilon} < 1 + \epsilon$.

To prove (13), the idea is the same as the first proof provided. One defines a random matrix and proceed using concentration inequalities. However, because the norms are squared, the middle term $\|A(x_i) - A(x_j)\|^2$ will not be distributed normally. It will have a chi-squared distribution which has its own concentration inequalities. In [5], the bound on the dimension is worse (larger), but the probability that a given map satisfies the lemma is at least $\frac{1}{M^2}$. This provides the additional claim that a map can be found in randomized polynomial time.

References

[1] Dmitry Panchenko, *Gaussian Distributions with Applications* unpublished.

[2] David Miyamoto, *Dvoretzky's Theorem and Concentration of Measure* unpublished.

[3] Yehoram Gordon, *Applications of the Gaussian Min-Max Theorem* unpublished.

[4] Sham Kakade, Greg Shakhnarovich. *Random Projections* CMSC 35900, 2009.

[5] Sanjoy Dasgupta, Anupam Gupta. *An Elementary Proof of the Theorem of Johnson and Lindenstrauss* 2002.

[6] William Johnson, Joram Lindenstrauss. *Extensions of Lipschitz mappings into a Hilbert space* Contemporary Mathematics, 1984.