Johnson-Lindenstrauss Lemma: For any set of points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ and $\epsilon > 0$ there exists a linear map $M : \mathbb{R}^d \to \mathbb{R}^m$ such that $m = O\left(\frac{\log n}{\epsilon^2}\right)$ and letting $\tilde{x}_i = M\mathbf{x}_i$:

For all $i, j$:

$$(1 - \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2.$$

Further, if $M \in \mathbb{R}^{m \times d}$ has each entry chosen independently from $\mathcal{N}(0, 1/m)$, it satisfies the guarantee with high probability.
Johnson-Lindenstrauss Lemma: For any set of points \( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \) and \( \epsilon > 0 \) there exists a linear map \( M : \mathbb{R}^d \rightarrow \mathbb{R}^m \) such that \( m = O \left( \frac{\log n}{\epsilon^2} \right) \) and letting \( \tilde{x}_i = M \vec{x}_i \):

\[
\text{For all } i, j : (1 - \epsilon) \| \vec{x}_i - \vec{x}_j \|_2 \leq \| \tilde{x}_i - \tilde{x}_j \|_2 \leq (1 + \epsilon) \| \vec{x}_i - \vec{x}_j \|_2.
\]

Further, if \( M \in \mathbb{R}^{m \times d} \) has each entry chosen independently from \( \mathcal{N}(0, 1/m) \), it satisfies the guarantee with high probability.

For \( d = 1 \text{ trillion}, \epsilon = .05, \) and \( n = 100,000, \) \( m \approx 6600. \)
Johnson-Lindenstrauss Lemma: For any set of points $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ and $\epsilon > 0$ there exists a linear map $M : \mathbb{R}^d \rightarrow \mathbb{R}^m$ such that $m = O\left(\frac{\log n}{\epsilon^2}\right)$ and letting $\tilde{x}_i = M\vec{x}_i$:

For all $i, j$:

$$(1 - \epsilon)\|\vec{x}_i - \vec{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\vec{x}_i - \vec{x}_j\|_2.$$ 

Further, if $M \in \mathbb{R}^{m \times d}$ has each entry chosen independently from $\mathcal{N}(0, 1/m)$, it satisfies the guarantee with high probability.

For $d = 1$ trillion, $\epsilon = .05$, and $n = 100,000$, $m \approx 6600$.

Very surprising! Powerful result with a simple construction: applying a random linear transformation to a set of points preserves distances between all those points with high probability.
For any $\tilde{x}_1, \ldots, \tilde{x}_n$ and $M \in \mathbb{R}^{m \times d}$ with each entry chosen independently from $\mathcal{N}(0, 1/m)$, with high probability, letting $\tilde{x}_i = M\tilde{x}_i$:

For all $i, j$:

$$(1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2.$$
For any $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{M} \in \mathbb{R}^{m \times d}$ with each entry chosen independently from $\mathcal{N}(0, 1/m)$, with high probability, letting $\tilde{\mathbf{x}}_i = \mathbf{M} \mathbf{x}_i$:

For all $i, j$: $(1 - \epsilon) \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2 \leq (1 + \epsilon) \|\mathbf{x}_i - \mathbf{x}_j\|_2$.

- $\mathbf{M}$ is known as a random projection. It is a random linear function, mapping length $d$ vectors to length $m$ vectors.
For any \( \vec{x}_1, \ldots, \vec{x}_n \) and \( M \in \mathbb{R}^{m \times d} \) with each entry chosen independently from \( \mathcal{N}(0, 1/m) \), with high probability, letting \( \tilde{x}_i = M\vec{x}_i \):

For all \( i, j \):

\[
(1 - \epsilon)\|\vec{x}_i - \vec{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\vec{x}_i - \vec{x}_j\|_2.
\]

- \( M \) is known as a random projection. It is a random linear function, mapping length \( d \) vectors to length \( m \) vectors.
- \( M \) is data oblivious. Stark contrast to methods like PCA.
• Alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured, etc. $\rightarrow$ efficient computation of $\tilde{x}_i = M\tilde{x}_i$.
• Alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured, etc. \( \implies \) efficient computation of \( \tilde{x}_i = M\tilde{x}_i. \)

• Data oblivious property means that once \( M \) is chosen, \( \tilde{x}_1, \ldots, \tilde{x}_n \) can be computed in a stream with little memory.

• Storage is just \( O(nm) \) rather than \( O(nd) \).
• Alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured, etc. $\Rightarrow$ efficient computation of $\tilde{x}_i = M\tilde{x}_i$.

• Data oblivious property means that once $M$ is chosen, $\tilde{x}_1, \ldots, \tilde{x}_n$ can be computed in a stream with little memory.

• Storage is just $O(nm)$ rather than $O(nd)$.

• Compression can be performed in parallel on different servers.
• Alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured, etc. \( \Rightarrow \) efficient computation of \( \tilde{x}_i = M\tilde{x}_i \).

• Data oblivious property means that once \( M \) is chosen, \( \tilde{x}_1, \ldots, \tilde{x}_n \) can be computed in a stream with little memory.

• Storage is just \( O(nm) \) rather than \( O(nd) \).

• Compression can be performed in parallel on different servers.

• When new data points are added, can be easily compressed, without updating existing points.
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $\vec{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

$$(1 - \epsilon) \| \vec{y} \|_2 \leq \| M\vec{y} \|_2 \leq (1 + \epsilon) \| \vec{y} \|_2$$

$M \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $\vec{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

$$
(1 - \epsilon)\|\vec{y}\|_2 \leq \|M\vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2
$$

I.e., applying a random matrix $M$ to any vector $\vec{y}$ preserves the norm with high probability. Like a low-distortion embedding, but for the length of a compressed vector rather than distances between vectors.

$M \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
**Distributional JL Lemma $\implies$ JL Lemma:** Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors.
Distributional JL Lemma $\Rightarrow$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

Proof: Given $x_1, \ldots, x_n$, define $(n \choose 2)$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$. 
**Distributional JL Lemma** $\implies$ **JL Lemma:** Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

**Proof:** Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2}\log 1/\delta'\right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

  $$
  (1 - \epsilon)\|y_{ij}\|_2 \leq \|My_{ij}\|_2 \leq (1 + \epsilon)\|y_{ij}\|_2
  $$


Distributional JL Lemma $\Rightarrow$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

Proof: Given $x_1, \ldots, x_n$, define $\left( \begin{array}{c} n \\ 2 \end{array} \right)$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O \left( \epsilon^{-2} \log 1/\delta' \right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

$$ (1 - \epsilon) \|y_{ij}\|_2 \leq \|My_{ij}\|_2 \leq (1 + \epsilon) \|y_{ij}\|_2 $$
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

**Proof:** Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2}\log 1/\delta'\right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

$$
(1 - \epsilon)\|x_i - x_j\|_2 \leq \|M(x_i - x_j)\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2
$$
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

**Proof:** Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2}\log 1/\delta'\right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

\[
(1 - \epsilon)\|x_i - x_j\|_2 \leq \|Mx_i - Mx_j\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2
\]
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

Proof: Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2}\log 1/\delta'\right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

\[
(1 - \epsilon)\|x_i - x_j\|_2 \leq \|Mx_i - Mx_j\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2
\]
**Distributional JL Lemma** $\implies$ **JL Lemma:** Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

**Proof:** Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2} \log 1/\delta’\right)$, for each $y_{ij}$ with probability at least $1 - \delta’$ we have:

$$
(1 - \epsilon)\|x_i - x_j\|_2 \leq \|M x_i - M x_j\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2
$$

- Union Bound: Every distance preserved with probability $1 - \binom{n}{2} \cdot \delta’$. 


Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $M$ preserves the norm of any $y$. The main JL Lemma says that $M$ preserves distances between vectors. Since $M$ is linear these are the same thing!

Proof: Given $x_1, \ldots, x_n$, define $\binom{n}{2}$ vectors $y_{ij}$ where $y_{ij} = x_i - x_j$.

- If we choose $M$ with $m = O\left(\epsilon^{-2}\log 1/\delta'\right)$, for each $y_{ij}$ with probability at least $1 - \delta'$ we have:

\[
(1 - \epsilon)\|x_i - x_j\|_2 \leq \|Mx_i - Mx_j\|_2 \leq (1 + \epsilon)\|x_i - x_j\|_2
\]

- Union Bound: Every distance preserved with probability $1 - \binom{n}{2} \cdot \delta'$.
- Setting $\delta' = \delta/\binom{n}{2}$ ensures all distances preserved with probability $1 - \delta$ and

\[
m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right) = O\left(\frac{\log(\binom{n}{2}/\delta)}{\epsilon^2}\right) = O\left(\frac{\log(n/\delta)}{\epsilon^2}\right)
\]
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

• Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle$
**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$. 
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon) \|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon) \|y\|_2.$$

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:

$$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^{d} \mathbb{E}[g_i] \cdot y_i = 0.$$
**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:

$$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^{d} \mathbb{E}[g_i] \cdot y_i = 0.$$

- By linearity of variance:

$$\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j] = \sum_{i=1}^{d} \text{Var}[g_i \cdot y_i] = \sum_{i} y_i^2 / m = \|y\|_2^2 / m.$$
DISTRIBUTIONAL JL PROOF

Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^d g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:
  $$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^d \mathbb{E}[g_i] \cdot y_i = 0.$$ 
- By linearity of variance:
  $$\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j] = \sum_{i=1}^d \text{Var}[g_i \cdot y_i] = \sum_i y_i^2 / m = \|y\|_2^2 / m.$$ 
- And hence $\mathbb{E}[\sum_j \tilde{y}_j^2] = \|y\|_2^2.$
Letting \( \tilde{y} = My \), we have \( \tilde{y}_j = \langle M_j, y \rangle \) and:

\[
\tilde{y}_j = \sum_{i=1}^{d} g_i \cdot y_i \text{ where } g_i \cdot y_i \sim \mathcal{N}(0, y_i^2/m).
\]

**Stability of Gaussian Random Variables.** For independent \( a \sim \mathcal{N}(\mu_1, \sigma_1^2) \) and \( b \sim \mathcal{N}(\mu_2, \sigma_2^2) \) we have:

\[
a + b \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)
\]

Thus, \( \tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \).