Johnson-Lindenstrass

Points: $x_1, \ldots, x_n \in \mathbb{R}^d$. 

Random $k = c \log n \varepsilon^2$ dimensional subspace. 

Claim: with probability $1 - \frac{1}{n^{c-2}}$,

\[
(1 - \varepsilon) \sqrt{k \cdot d} |x_i - x_j|^2 \leq |y_i - y_j|^2 \leq (1 + \varepsilon) \sqrt{k \cdot d} |x_i - x_j|^2
\]

"Projecting and scaling by $\sqrt{d \cdot k}$ preserves all pairwise distances within factor of $1 \pm \varepsilon$."
Johnson-Lindenstrass

Points: $x_1, \ldots, x_n \in \mathbb{R}^d$.

Random $k = \frac{c \log n}{\varepsilon^2}$ dimensional subspace.
Points: $x_1, \ldots, x_n \in \mathbb{R}^d$.

Random $k = \frac{c \log n}{\epsilon^2}$ dimensional subspace.

Claim: with probability $1 - \frac{1}{n^{c-2}}$,

$$(1 - \epsilon) \sqrt{\frac{k}{d}} |x_i - x_j|^2 \leq |y_i - y_j|^2 \leq (1 + \epsilon) \sqrt{\frac{k}{d}} |x_i - x_j|^2$$
Points: $x_1, \ldots, x_n \in \mathbb{R}^d$.

Random $k = \frac{c \log n}{\varepsilon^2}$ dimensional subspace.

Claim: with probability $1 - \frac{1}{n^{c-2}}$,

$$(1 - \varepsilon) \sqrt{\frac{k}{d}} |x_i - x_j|^2 \leq |y_i - y_j|^2 \leq (1 + \varepsilon) \sqrt{\frac{k}{d}} |x_i - x_j|^2$$

“Projecting and scaling by $\sqrt{\frac{d}{k}}$ preserves all pairwise distances w/in factor of $1 \pm \varepsilon$.”
Random subspace.

Method 1:
Random subspace.

Method 1:
Pick unit $v_1$
Random subspace.

Method 1:
Pick unit $v_1$, 

Method 2:
Choose $k$ vectors $v_1,...,v_k$ Gram Schmidt orthonormalization of $k \times d$ matrix where rows are $v_i$.
remove projection onto previous subspace.
Random subspace.

Method 1:
Pick unit $v_1$,
$\nu_2$ orthogonal to $v_1$, 

...
Method 1:
Pick unit $v_1$,
$v_2$ orthogonal to $v_1$,
...
Random subspace.

Method 1:
Pick unit $\nu_1$, 
$\nu_2$ orthogonal to $\nu_1$, 
... 
$\nu_k$ orthogonal to previous vectors...
Random subspace.

Method 1:
Pick unit \( v_1 \),
\( v_2 \) orthogonal to \( v_1 \),
\[ \ldots \]
\( v_k \) orthogonal to previous vectors...

Method 2:
Method 1:
Pick unit $v_1$, 
$v_2$ orthogonal to $v_1$, 
... 
$v_k$ orthogonal to previous vectors...

Method 2:
Choose $k$ vectors $v_1, \ldots, v_k$
Random subspace.

Method 1:
Pick unit $v_1$,  
$v_2$ orthogonal to $v_1$,  
$\ldots$  
$v_k$ orthogonal to previous vectors...

Method 2:
Choose $k$ vectors $v_1, \ldots, v_k$  
Gram Schmidt orthonormalization of $k \times d$ matrix where rows are $v_i$. 
Random subspace.

Method 1:
Pick unit $v_1$,
$v_2$ orthogonal to $v_1$,
...
$v_k$ orthogonal to previous vectors...

Method 2:
Choose $k$ vectors $v_1, \ldots, v_k$
Gram Schmidt orthonormalization of $k \times d$ matrix where rows are $v_i$.
remove projection onto previous subspace.
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$. 
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$
Projections.

Project \( x \) into subspace spanned by \( v_1, v_2, \ldots, v_k \).

\[
y_1 = x \cdot v_1, \quad y_2 = x \cdot v_2, \ldots, \quad y_k = x \cdot v_k
\]

Projection: \( (y_1, \ldots, y_k) \).
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \ldots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \ldots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$. 

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$

Inverse of $U$ maps $e_i$ to random vector $v_i$ and $U^{-1} = U$. 

$z = Ux$ is uniformly distributed on $d$ sphere for unit $x \in \mathbb{R}^d$.

$y_i$ is $i$th coordinate of random vector $z$. 
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \ldots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \ldots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \ldots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \ldots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.
View As: Random vector, standard basis for $k$ dimensions.
Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle U v_i | U x \rangle$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$

Inverse of $U$ maps $e_i$ to random vector $v_i$
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \cdots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \cdots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$

Inverse of $U$ maps $e_i$ to random vector $v_i$ and $U^{-1} = U$. 

$z = Ux$ is uniformly distributed on $d$ sphere for unit $x \in \mathbb{R}^d$.

$y_i$ is $i$th coordinate of random vector $z$. 
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \ldots, v_k$.

$$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \ldots, y_k = x \cdot v_k$$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$$

Inverse of $U$ maps $e_i$ to random vector $v_i$ and $U^{-1} = U$.

$z = Ux$ is uniformly distributed on $d$ sphere for unit $x \in \mathbb{R}^d$. 
Projections.

Project $x$ into subspace spanned by $v_1, v_2, \ldots, v_k$.

$y_1 = x \cdot v_1, y_2 = x \cdot v_2, \ldots, y_k = x \cdot v_k$

Projection: $(y_1, \ldots, y_k)$.

Have: Arbitrary vector, random $k$-dimensional subspace.

View As: Random vector, standard basis for $k$ dimensions.

Orthogonal $U$ - rotates $v_1, \ldots, v_k$ onto $e_1, \ldots, e_k$

$y_i = \langle v_i | x \rangle = \langle Uv_i | Ux \rangle = \langle e_i | Ux \rangle = \langle e_i | z \rangle$

Inverse of $U$ maps $e_i$ to random vector $v_i$ and $U^{-1} = U$.

$z = Ux$ is uniformly distributed on $d$ sphere for unit $x \in \mathbb{R}^d$.

$y_i$ is $i$th coordinate of random vector $z$. 
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$. 

$\mathbb{E}\left[ \sum_{i \in [d]} z_i^2 \right] = 1.$

Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$\mathbb{E}\left[ \sum_{i \in [k]} z_i^2 \right] = k d.$

Linearity of Expectation.

Expected length is $\sqrt{kd}$.

Johnson-Lindenstrass: close to expectation.

$k$ is large enough $\rightarrow \approx (1 \pm \varepsilon) \sqrt{kd}$ with decent probability.
Expected value of $y_i$. 

Random projection: first $k$ coordinates of random unit vector, $z_i$. 

$E[\sum_{i \in [d]} z_i^2] = 1$. 

Linearity of Expectation. 

By symmetry, each $z_i$ is identically distributed. 

$E[\sum_{i \in [k]} z_i^2] = k$. 

Linearity of Expectation. 

Expected length is $\sqrt{k}d$. 

Johnson-Lindenstrass: close to expectation. $k$ is large enough $\to \approx (1 \pm \varepsilon)\sqrt{k}d$ with decent probability.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$$E[\sum_{i \in [d]} z_i^2] = 1.$$ Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. Linearity of Expectation.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. Linearity of Expectation.

Expected length is $\sqrt{\frac{k}{d}}$. 
Random projection: first $k$ coordinates of random unit vector, $z_i$. 

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. Linearity of Expectation.

Expected length is $\sqrt{\frac{k}{d}}$.

Johnson-Lindenstrass: close to expectation.
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. Linearity of Expectation.

Expected length is $\sqrt{\frac{k}{d}}$.

Johnson-Lindenstrass: close to expectation.

$k$ is large enough $\rightarrow$
Expected value of $y_i$.

Random projection: first $k$ coordinates of random unit vector, $z_i$.

$E[\sum_{i \in [d]} z_i^2] = 1$. Linearity of Expectation.

By symmetry, each $z_i$ is identically distributed.

$E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. Linearity of Expectation.

Expected length is $\sqrt{\frac{k}{d}}$.

Johnson-Lindenstrass: close to expectation.

$k$ is large enough $\rightarrow$

$\approx (1 \pm \epsilon)\sqrt{\frac{k}{d}}$ with decent probability.
Concentration Bounds.

$z$ is uniformly random unit vector.
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$. 

$\text{Claim: } P(|z_1| > \sqrt{\frac{d}{2}}) \leq e^{-\frac{t^2}{2}}$.

Sphere view: surface "far" from equator defined by $e_1$.

$\Delta |z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere.

Point on "$\Delta$-spherical cap".

Area of caps $\leq \text{S.A. of sphere of radius } \sqrt{1 - \Delta^2} 
\propto r^d = (1 - \Delta^2)^{d/2} \approx e^{-t^2d}$. 

Constant of $\propto$ is unit sphere area.

$P[\text{any } z_i > \sqrt{2\log d} E[z_i^2]]$ is small.
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$
Concentration Bounds.

$z$ is uniformly random unit vector.

Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$. 

Area of caps $\leq S.A.$ of sphere of radius $\sqrt{1 - \Delta^2}$

$\propto r^d = (1 - \Delta^2)^{d/2} = (1 - t^2/d)^{d/2} \approx e^{-t^2/2}$

Constant of $\propto$ is unit sphere area.
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$. 

Diagram: Sphere with a blue point labeled $\Delta$ at a distance $\Delta$ from the equator.
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if

$z \geq \Delta$ from equator of sphere.
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere.
Point on “$\Delta$-spherical cap”.

\[\text{Area of caps} \leq \text{S.A. of sphere of radius } \sqrt{1 - \Delta^2} \propto \frac{r_d}{d} = \frac{(1 - \Delta^2)^{d/2}}{d/2} \approx e^{-t^2/2} \text{d} \]

Constant of $\propto$ is unit sphere area.
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if
$z \geq \Delta$ from equator of sphere.
Point on “$\Delta$-spherical cap”.

Area of caps
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere.

Point on “$\Delta$-spherical cap”.

Area of caps $\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$.
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}.$

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1.$

$|z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere.

Point on “$\Delta$-spherical cap”.

Area of caps

$\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$

$\propto r^d = (1 - \Delta^2)^{d/2}$
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere.
Point on “$\Delta$-spherical cap”.

Area of caps

$\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$

$\propto r^d = (1 - \Delta^2)^{d/2}$

$\propto \left(1 - \frac{t^2}{d}\right)^{d/2}$
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if

$z \geq \Delta$ from equator of sphere.

Point on “$\Delta$-spherical cap”.

Area of caps

$\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$

$\propto r^d = (1 - \Delta^2)^{d/2}$

$\propto (1 - \frac{t^2}{d})^{d/2} \approx e^{-t^2/2} 2d$
Concentration Bounds.

$z$ is uniformly random unit vector.
Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if
$z \geq \Delta$ from equator of sphere.
Point on “$\Delta$-spherical cap”.

Area of caps
$\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$
$\propto r^d = (1 - \Delta^2)^{d/2}$
$\propto \left(1 - \frac{t^2}{d}\right)^{d/2} \approx e^{-t^2} 2d$

Constant of $\propto$ is unit sphere area.
Concentration Bounds.

$z$ is uniformly random unit vector. Random point on the unit sphere. $E[\sum_{i \in [k]} z_i^2] = \frac{k}{d}$.

Claim: $\Pr[|z_1| > \frac{t}{\sqrt{d}}] \leq e^{-t^2/2}$

Sphere view: surface “far” from equator defined by $e_1$.

$|z_1| \geq \Delta$ if $z \geq \Delta$ from equator of sphere. Point on “$\Delta$-spherical cap”.

Area of caps

$\leq$ S.A. of sphere of radius $\sqrt{1 - \Delta^2}$

$\propto r^d = (1 - \Delta^2)^{d/2}$

$\propto (1 - \frac{t^2}{d})^{d/2} \approx e^{-t^2/2} \frac{d}{2}$

Constant of $\propto$ is unit sphere area.

$\Pr[\text{any } z_i^2 > \sqrt{2\log d} E[z_i^2]]$ is small.
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2\log dE[z_i^2]}] \) is small.
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length?
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots + z_k^2 \).
Many coordinates.

Proved $\Pr[\text{any } z_i^2 > \sqrt{2 \log d} E[z_i^2]]$ is small.

Length? $z = z_1^2 + z_2^2 + \cdots z_k^2$.

$$\Pr\left[\left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}$$
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log d E[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2 d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log d} E[z_i^2]] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2 d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| \right]
\]
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log d E[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots + z_k^2 \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right]
\]
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2\log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots + z_k^2. \)

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \epsilon \sqrt{\frac{k}{d}}, \quad k = \frac{c\log n}{\epsilon^2}. \)

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \epsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\epsilon^2k}
\]
Many coordinates.

Proved $\Pr[\text{any } z_i^2 > \sqrt{2\log dE[z_i^2]}]$ is small.

Length? $z = z_1^2 + z_2^2 + \cdots z_k^2$.

$$\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}$$

Substituting $t = \varepsilon \sqrt{\frac{k}{d}}$, $k = \frac{c\log n}{\varepsilon^2}$.

$$\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2k} = e^{-c\log n} = \frac{1}{n^c}$$
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}}, k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2k} = e^{-c \log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \varepsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.
Many coordinates.

Proved $\Pr[\text{any } z_i^2 > \sqrt{2 \log d E[z_i^2]}]$ is small.

Length? $z = z_1^2 + z_2^2 + \cdots z_k^2$.

$$\Pr[\sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} > t] \leq e^{-t^2 d}$$

Substituting $t = \varepsilon \sqrt{\frac{k}{d}}$, $k = \frac{c \log n}{\varepsilon^2}$.

$$\Pr[\sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \varepsilon \sqrt{\frac{k}{d}}] \leq e^{-\varepsilon^2 k} = e^{-c \log n} = \frac{1}{n^c}$$

**Johnson-Lindenstrauss:** For $n$ points, $x_1, \ldots, x_n$, all distances preserved to within $1 \pm \varepsilon$ under $\sqrt{\frac{k}{d}}$-scaled projection above.

View one pair $x_i - x_j$ as vector.
Many coordinates.

Proved $\Pr[\text{any } z_i^2 > \sqrt{2\log dE[z_i^2]}]$ is small.

Length? $z = z_1^2 + z_2^2 + \cdots + z_k^2$.

$$\Pr[\left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t] \leq e^{-t^2d}$$

Substituting $t = \varepsilon \sqrt{\frac{k}{d}}$, $k = \frac{c\log n}{\varepsilon^2}$.

$$\Pr[\left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}}] \leq e^{-\varepsilon^2k} = e^{-c\log n} = \frac{1}{n^c}$$

**Johnson-Lindenstrauss:** For $n$ points, $x_1, \ldots, x_n$, all distances preserved to within $1 \pm \varepsilon$ under $\sqrt{\frac{k}{d}}$-scaled projection above.

View one pair $x_i - x_j$ as vector.
Scale to unit.
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr\left[ \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} > t \right] \leq e^{-t^2 d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}}, k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2 k} = e^{-c \log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \varepsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.

View one pair \( x_i - x_j \) as vector.

Scale to unit.

Projection fails to preserve \( |x_i - x_j| \)
Many coordinates.
Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots + z_k^2. \)

\[
\Pr[\left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t] \leq e^{-t^2d}
\]

Substituting \( t = \epsilon \sqrt{\frac{k}{d}}, k = \frac{c \log n}{\epsilon^2}. \)

\[
\Pr[\left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \epsilon \sqrt{\frac{k}{d}}] \leq e^{-\epsilon^2k} = e^{-c \log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \epsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.

View one pair \( x_i - x_j \) as vector.
Scale to unit.
Projection fails to preserve \( |x_i - x_j| \)
with probability \( \leq \frac{1}{n^c} \)
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2 \log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2. \)

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2k} = e^{-c \log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \varepsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.

View one pair \( x_i - x_j \) as vector.

Scale to unit.

Projection fails to preserve \( |x_i - x_j| \)

with probability \( \leq \frac{1}{n^c} \)

Scaled vector length also preserved.
Many coordinates.
Proved $\Pr[\text{any } z_i^2 > \sqrt{2\log d}E[z_i^2]]$ is small.

Length? $z = z_1^2 + z_2^2 + \cdots z_k^2$.

$$\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}$$

Substituting $t = \varepsilon\sqrt{\frac{k}{d}}$, $k = \frac{c\log n}{\varepsilon^2}$.

$$\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon\sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2 k} = e^{-c\log n} = \frac{1}{n^c}$$

**Johnson-Lindenstrauss:** For $n$ points, $x_1, \ldots, x_n$, all distances preserved to within $1 \pm \varepsilon$ under $\sqrt{\frac{k}{d}}$-scaled projection above.

View one pair $x_i - x_j$ as vector.
Scale to unit.
Projection fails to preserve $|x_i - x_j|$ with probability $\leq \frac{1}{n^c}$
Scaled vector length also preserved.

$\leq n^2$ pairs
Many coordinates.

Proved Pr[any \( z_i^2 > \sqrt{2 \log dE[z_i^2]} \)] is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr\left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2 k} = e^{-c \log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \varepsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.

View one pair \( x_i - x_j \) as vector.

Scale to unit.

Projection fails to preserve \( |x_i - x_j| \)

with probability \( \leq \frac{1}{n^c} \)

Scaled vector length also preserved.

\( \leq n^2 \) pairs plus union bound
Many coordinates.

Proved \( \Pr[\text{any } z_i^2 > \sqrt{2\log dE[z_i^2]}] \) is small.

Length? \( z = z_1^2 + z_2^2 + \cdots z_k^2 \).

\[
\Pr \left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > t \right] \leq e^{-t^2d}
\]

Substituting \( t = \varepsilon \sqrt{\frac{k}{d}} \), \( k = \frac{c \log n}{\varepsilon^2} \).

\[
\Pr \left[ \left| \sqrt{z_1^2 + z_2^2 + \cdots + z_k^2} - \sqrt{\frac{k}{d}} \right| > \varepsilon \sqrt{\frac{k}{d}} \right] \leq e^{-\varepsilon^2k} = e^{-c\log n} = \frac{1}{n^c}
\]

**Johnson-Lindenstrauss:** For \( n \) points, \( x_1, \ldots, x_n \), all distances preserved to within \( 1 \pm \varepsilon \) under \( \sqrt{\frac{k}{d}} \)-scaled projection above.

View one pair \( x_i - x_j \) as vector.
Scale to unit.
Projection fails to preserve \( |x_i - x_j| \)

with probability \( \leq \frac{1}{n^c} \)

Scaled vector length also preserved.

\( \leq n^2 \) pairs plus union bound

\( \rightarrow \) prob any pair fails to be preserved with \( \leq \frac{1}{n^{c-2}} \).
Locality Preserving Hashing

Find nearby points in high dimensional space.
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$. 
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation.
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation. Not quite a solution.
Locality Preserving Hashing

Find nearby points in high dimensional space.
Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation.
Not quite a solution. Why?
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation. Not quite a solution. Why?

Close to grid boundary.
Locality Preserving Hashing

Find nearby points in high dimensional space.
  Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation.
Not quite a solution. Why?
  Close to grid boundary.
Find close points to $x$: 
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation. Not quite a solution. Why?
   Close to grid boundary.

Find close points to $x$:
   Check grid cell and neighboring grid cells.
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation. Not quite a solution. Why?
  Close to grid boundary.
Find close points to $x$:
  Check grid cell and neighboring grid cells.
Locality Preserving Hashing

Find nearby points in high dimensional space. Points could be images!

Hash function $h(\cdot)$ s.t. $h(x_i) = h(x_j)$ if $d(x_i, x_j) \leq \delta$.

Low dimensions: grid cells give $\sqrt{d}$-approximation. Not quite a solution. Why?
   Close to grid boundary.
Find close points to $x$:
   Check grid cell and neighboring grid cells.

Project high dimensional points into low dimensions.
Locality Preserving Hashing

Find nearby points in high dimensional space.
Points could be images!

Hash function \( h(\cdot) \) s.t. \( h(x_i) = h(x_j) \) if \( d(x_i, x_j) \leq \delta \).

Low dimensions: grid cells give \( \sqrt{d} \)-approximation.
Not quite a solution. Why?
Close to grid boundary.
Find close points to \( x \):
Check grid cell and neighboring grid cells.

Project high dimensional points into low dimensions.

Use grid hash function.
Implementing Johnson-Lindenstrauss

Random vectors
Implementing Johnson-Lindenstrauss

Random vectors have many bits
Random vectors have many bits
Use random bit vectors: \(\{-1, +1\}^d\) instead.
Implementing Johnson-Lindenstrauss

Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.
   Almost orthogonal.
Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.
   Almost orthogonal.
Project z.
Implementing Johnson-Lindenstrauss

Random vectors have many bits
Use random bit vectors: \( \{-1, +1\}^d \) instead.
  Almost orthogonal.
Project \( z \).
Coordinate for bit vector \( b \).
Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.
Almost orthogonal.

Project $z$.

Coordinate for bit vector $b$.

$$C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i$$
Random vectors have many bits
Use random bit vectors: \([-1, +1]^d\) instead.
   Almost orthogonal.
Project \(z\).
Coordinate for bit vector \(b\).
   \[ C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i \]
   \[ E[C_i^2] = \]
Random vectors have many bits

Use random bit vectors: $\{-1, +1\}^d$ instead.

Almost orthogonal.

Project $z$.

Coordinate for bit vector $b$.

\[ C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i \]

\[ E[C_i^2] = E[\frac{1}{d} \sum_{i,j} b_i b_j z_i z_j] \]
Implementing Johnson-Lindenstrauss

Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.

Almost orthogonal.

Project $z$.

Coordinate for bit vector $b$.

$$C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i$$

$$E[C_i^2] = E[\frac{1}{d} \sum_{i,j} b_i b_j z_i z_j] = \frac{1}{d} \sum_{i,j} E[b_i b_j] z_i z_j = \frac{1}{d} \sum_i z_i^2$$
Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.

Almost orthogonal.

Project $z$.

Coordinate for bit vector $b$.

$$C_i = \frac{1}{\sqrt{d}} \sum b_i z_i$$

$$E[C_i^2] = E[\frac{1}{d} \sum_{i,j} b_i b_j z_i z_j] = \frac{1}{d} \sum_{i,j} E[b_i b_j] z_i z_j = \frac{1}{d} \sum z_i^2 = \frac{1}{d}$$
Random vectors have many bits
Use random bit vectors: $\{-1, +1\}^d$ instead.
Almost orthogonal.
Project $z$.
Coordinate for bit vector $b$.
$$C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i$$
$$E[C_i^2] = E[\frac{1}{d} \sum_{i,j} b_i b_j z_i z_j] = \frac{1}{d} \sum_{i,j} E[b_i b_j] z_i z_j = \frac{1}{d} \sum_i z_i^2 = \frac{1}{d}$$
Implementing Johnson-Lindenstrauss

Random vectors have many bits
Use random bit vectors: \([-1, +1]^d\) instead.

Almost orthogonal.

Project \(z\).

Coordinate for bit vector \(b\).

\[
C_i = \frac{1}{\sqrt{d}} \sum_i b_i z_i
\]

\[
E[C_i^2] = E\left[\frac{1}{d} \sum_{i,j} b_i b_j z_i z_j\right] = \frac{1}{d} \sum_{i,j} E[b_i b_j] z_i z_j = \frac{1}{d} \sum_i z_i^2 = \frac{1}{d}
\]

\[
E[\sum_i C_i^2] = \frac{k}{d}
\]
Binary Johnson-Lindenstrass

Project onto $[-1, +1]$ vectors.
Binary Johnson-Lindenstrass

Project onto $[-1, +1]$ vectors.

$$E[C] = E[\sum_i C_i^2] = \frac{k}{d}$$
Binary Johnson-Lindenstrass

Project onto $[-1, +1]$ vectors.

$E[C] = E[\sum_i C_i^2] = \frac{k}{d}$

Concentration?
Binary Johnson-Lindenstrass

Project onto $[-1, +1]$ vectors.

$$E[C] = E[\sum_i C_i^2] = \frac{k}{d}$$

Concentration?

$$\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$
Project onto $[-1, +1]$ vectors.

$$E[C] = E[\sum_i C_i^2] = \frac{k}{d}$$

Concentration?

$$\Pr \left[ |C - \frac{k}{d}| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$

Choose $k = \frac{c \log n}{\varepsilon^2}$. 
Binary Johnson-Lindenstrass

Project onto $[-1, +1]$ vectors.

$$E[C] = E[\sum_i C_i^2] = \frac{k}{d}$$

Concentration?

$$\Pr\left[|C - \frac{k}{d}| \geq \varepsilon \frac{k}{d}\right] \leq e^{-\varepsilon^2 k}$$

Choose $k = \frac{c \log n}{\varepsilon^2}$.

$\rightarrow$ failure probability $\leq 1/n^c$. 
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]
Analysis Idea.

\[
\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}
\]

Variance of \( C_i^2 \)?
Analysis Idea.

$$\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$

Variance of $C_i^2$? \[ \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \]
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]

Variance of \( C_i^2 \)? \( \left( \frac{k}{d^2} \right) \left( \sum i z_i^4 + 4 \sum i, j z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum z_i^2)^2 \)
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]

Variance of \( C_i^2 \)? \( \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2 \left( \sum_i z_i^2 \right)^2 \leq \frac{2k}{d^2} \).
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]

Variance of \( C_i^2 \)? \( \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2 \left( \sum_i z_i^2 \right)^2 \leq \frac{2k}{d^2} \).

Roughly normal (gaussian):
Analysis Idea.

\[
Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}
\]

Variance of \( C_i^2 \)? \( \left( \frac{k}{d^2} \right) ( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 ) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2} \).

Roughly normal (gaussian):

Density \( \propto e^{-t^2}/2 \) for \( t \) std deviations away.
Analysis Idea.

\[
\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}
\]

Variance of \( C_i^2 \)?

\[
\left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2 \left( \sum_i z_i^2 \right)^2 \leq \frac{2k}{d^2}.
\]

Roughly normal (gaussian):

Density \( \propto e^{-t^2}/2 \) for \( t \) std deviations away.

So, assuming normality
Analysis Idea.

$$\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$

Variance of $C_i^2$? \( \left( \frac{k}{d^2} \right) (\sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2}. \)

Roughly normal (gaussian):

- Density $\propto e^{-t^2/2}$ for $t$ std deviations away.

So, assuming normality

$$\sigma = \frac{\sqrt{k}}{d},$$
Analysis Idea.

$$\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$

Variance of $C_i^2$: \( \left( \frac{k}{d^2} \right) \left( \sum i z_i^4 + 4 \sum i,j z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum i z_i^2)^2 \leq \frac{2k}{d^2} \).

Roughly normal (gaussian):

Density $\propto e^{-t^2}/2$ for $t$ std deviations away.

So, assuming normality

$$\sigma = \frac{\sqrt{k}}{d}, \ t = \frac{\varepsilon k}{\sqrt{2k} d}$$
Analysis Idea.

\[
\Pr \left[ \left| C - \frac{k}{d} \right| \geq \epsilon \frac{k}{d} \right] \leq e^{-\epsilon^2 k}
\]

Variance of \(C_i^2\)? \[\left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2}.
\]

Roughly normal (gaussian):
Density \(\propto e^{-t^2/2}\) for \(t\) std deviations away.

So, assuming normality
\[
\sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\epsilon \frac{k}{d}}{\sqrt{2k}} = \epsilon \frac{\sqrt{k}}{\sqrt{2}}.
\]
Analysis Idea.

\[
\Pr \left[ \left| C - \frac{k}{d} \right| \geq \epsilon \frac{k}{d} \right] \leq e^{-\epsilon^2 k}
\]

Variance of \( C_i^2 \)? \((\frac{k}{d^2}) (\sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2) \leq \left(\frac{k}{d^2}\right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2}.

Roughly normal (gaussian):
Density \( \propto e^{-t^2/2} \) for \( t \) std deviations away.

So, assuming normality
\[
\sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\epsilon k}{\sqrt{2kd}} = \epsilon \sqrt{k}/\sqrt{2}.
\]

Probability of failure roughly \( \leq e^{-t^2/2} \)
Analysis Idea.

\[
\Pr \left[ \left| C - \frac{kd}{d} \right| \geq \varepsilon \frac{kd}{d} \right] \leq e^{-\varepsilon^2 k}
\]

Variance of \( C_i \)? \( \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2} \).

Roughly normal (gaussian):

Density \( \propto e^{-t^2/2} \) for \( t \) std deviations away.

So, assuming normality

\[
\sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\varepsilon \frac{k}{d}}{\sqrt{2k}} = \varepsilon \sqrt{k}/\sqrt{2}.
\]

Probability of failure roughly \( \leq e^{-t^2/2} \)

\( \rightarrow e^{\varepsilon^2 k/4} \)
Analysis Idea.

$$\Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k}$$

Variance of $C_i^2$? $\left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2}$. 

Roughly normal (gaussian):

Density $\propto e^{-t^2/2}$ for $t$ std deviations away.

So, assuming normality

$$\sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\varepsilon \frac{k}{d}}{\sqrt{2\frac{k}{d}}} = \varepsilon \sqrt{k}/\sqrt{2}.$$ 

Probability of failure roughly $\leq e^{-t^2/2}$

$\rightarrow e^{\varepsilon^2 k/4}$

“Roughly normal.”
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]

Variance of \( C_i^2 \)? \( \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2(\sum_i z_i^2)^2 \leq \frac{2k}{d^2} \).

Roughly normal (gaussian):
Density \( \propto e^{-t^2/2} \) for \( t \) std deviations away.

So, assuming normality
\[ \sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\varepsilon \frac{k}{d}}{\sqrt{2k/d}} = \varepsilon \sqrt{k/\sqrt{2}}. \]

Probability of failure roughly \( \leq e^{-t^2/2} \)
\[ \rightarrow e^{\varepsilon^2 k/4} \]

“Roughly normal.” Chernoff, Berry-Esseen, Central Limit Theorems.
Analysis Idea.

\[ \Pr \left[ \left| C - \frac{k}{d} \right| \geq \varepsilon \frac{k}{d} \right] \leq e^{-\varepsilon^2 k} \]

Variance of \( C_i^2 \)?
\[ \left( \frac{k}{d^2} \right) \left( \sum_i z_i^4 + 4 \sum_{i,j} z_i^2 z_j^2 \right) \leq \left( \frac{k}{d^2} \right) 2 \left( \sum_i z_i^2 \right)^2 \leq \frac{2k}{d^2}. \]

Roughly normal (gaussian):
Density \( \propto e^{-t^2}/2 \) for \( t \) std deviations away.

So, assuming normality
\[ \sigma = \frac{\sqrt{k}}{d}, \quad t = \frac{\varepsilon \frac{k}{d}}{\sqrt{2k} \frac{d}{d}} = \varepsilon \sqrt{k}/\sqrt{2}. \]

Probability of failure roughly \( \leq e^{-t^2}/2 \)
\[ \rightarrow e^{\varepsilon^2 k/4} \]

“Roughly normal.” Chernoff, Berry-Esseen, Central Limit Theorems.
Sum up
Have a good break!