

Random Projections

2IMW30 - Foundations of data mining
TU Eindhoven, Quartile 3, 2016-2017

Anne Driemel

Why reduce the dimension?

Representation of input data often is often high dimensional (images, documents, etc.)

There are two main reasons to reduce the dimension:

- some algorithms have **running time** exponential in the dimension
- we want to **visualize** inherent structure in the data

Why reduce the dimension?

Representation of input data often is often high dimensional (images, documents, etc.)

There are two main reasons to reduce the dimension:

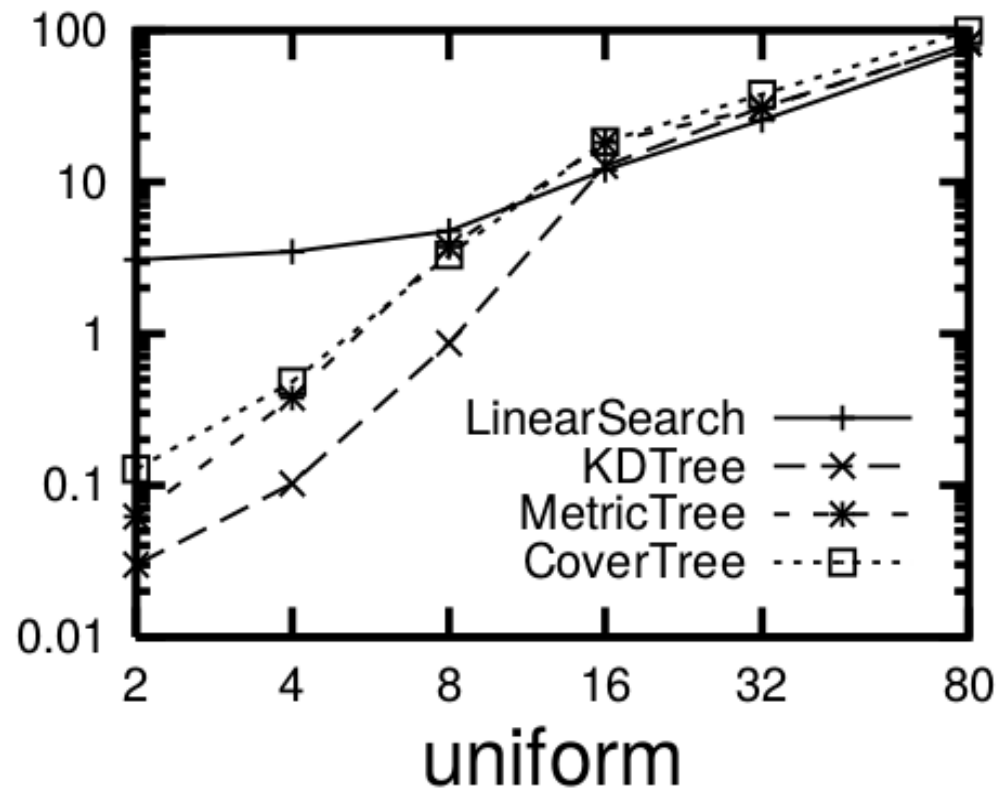
- some algorithms have **running time** exponential in the dimension
- we want to **visualize** inherent structure in the data

Overview of this lecture

- Nearest-neighbor searching
- Embedding and Distortion
- Achlioptas' Random Projection
- Projection onto a subspace
- Random Rotation (Expectation)
- Analysis of a fixed distance (Expectation)
- Law of large numbers
- Concentration of measure
- Analysis of a fixed distance
- Analysis of the Distortion
- Alternative Projection Matrix

Nearest neighbor searching

CPU-time to query the k -nearest neighbors vs. dimension of the data

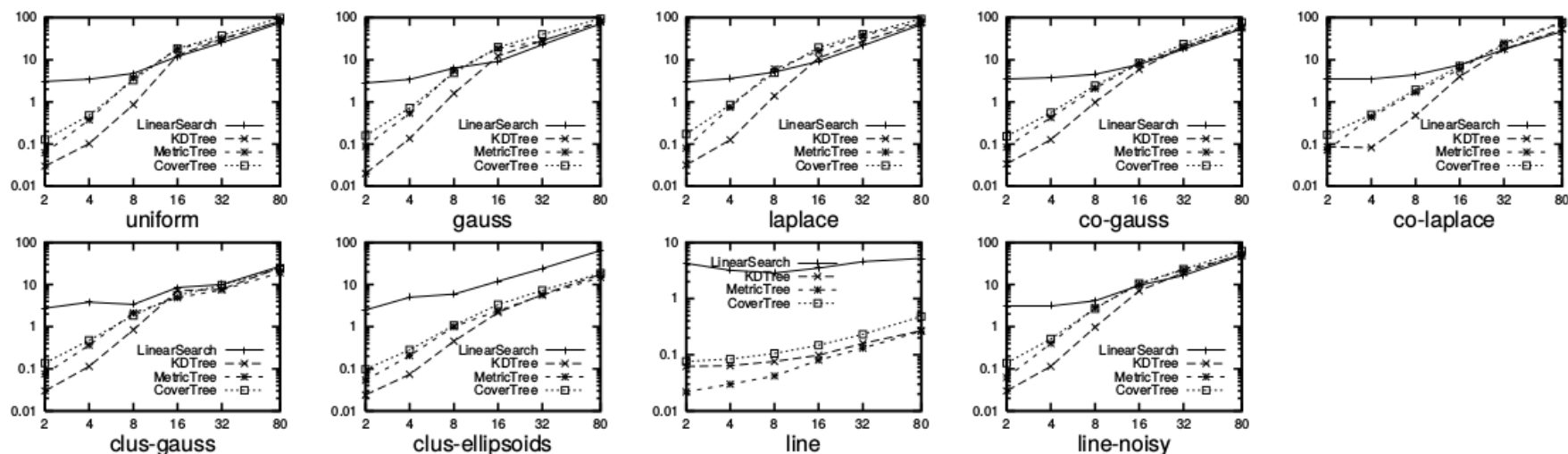


Source: Ashraf M. Kibriya and Eibe Frank "An Empirical Comparison of Exact Nearest Neighbour Algorithms" PKDD 2007

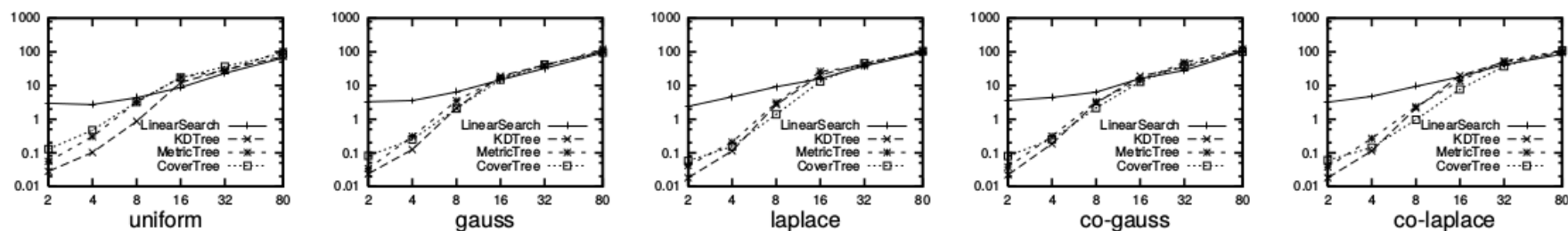
Nearest neighbor searching

CPU-time to query the k -nearest neighbors vs. dimension of the data

CPUQueryTime vs Dim (K=5 n=16000)
Non-uniform Query



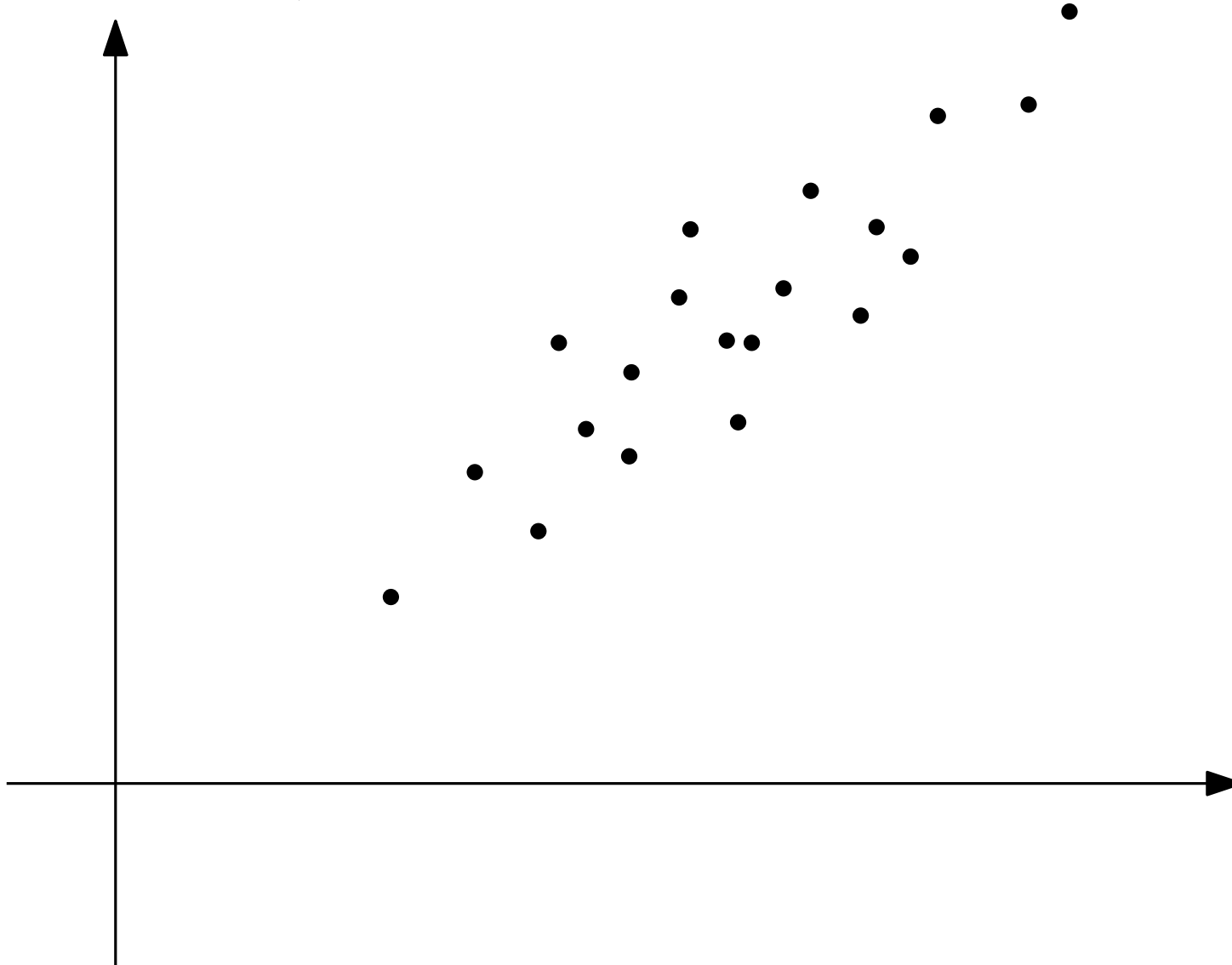
Uniform Query



Source: Ashraf M. Kibriya and Eibe Frank "An Empirical Comparison of Exact Nearest Neighbour Algorithms" PKDD 2007

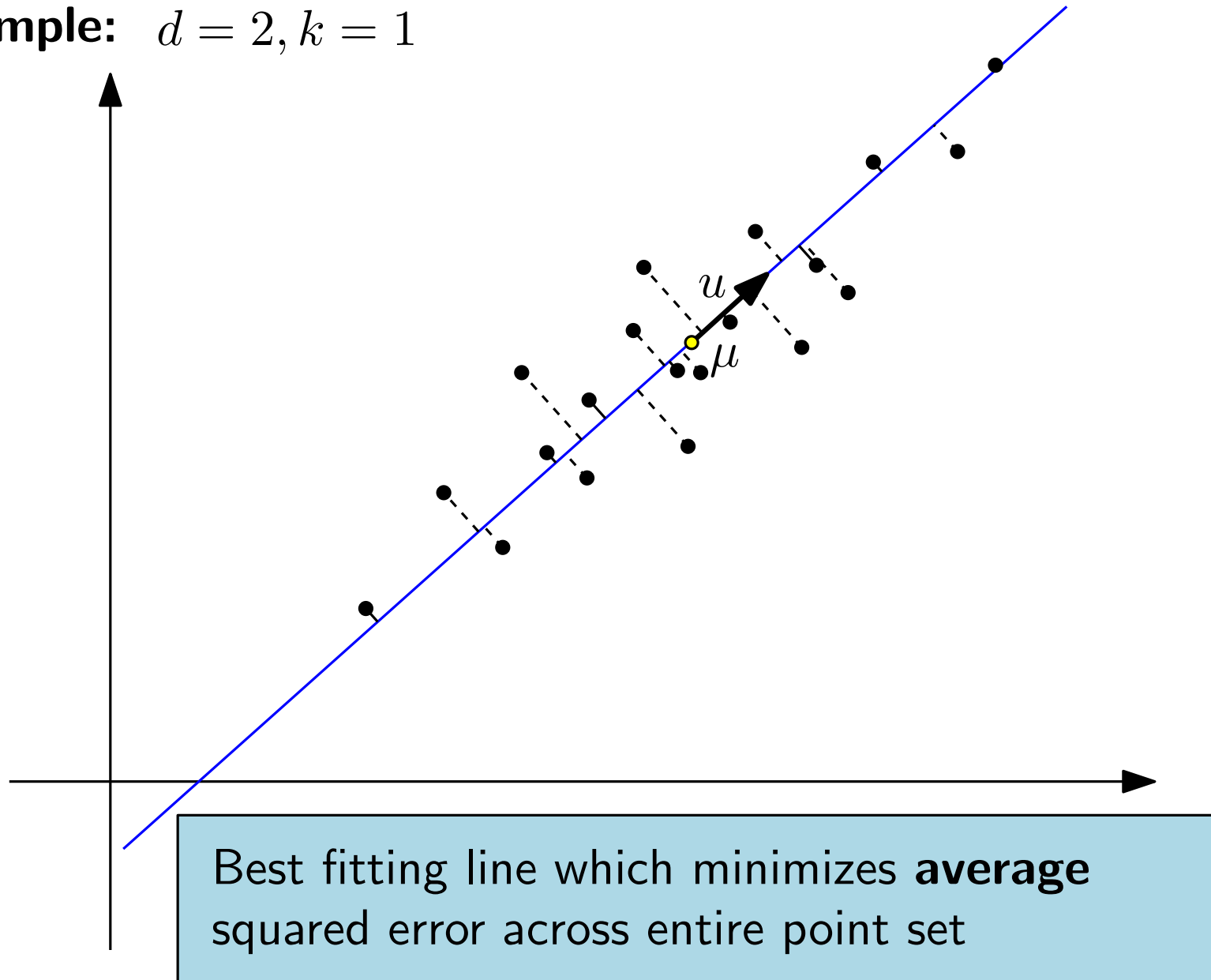
Principal Component Analysis (PCA)

Example: $d = 2, k = 1$



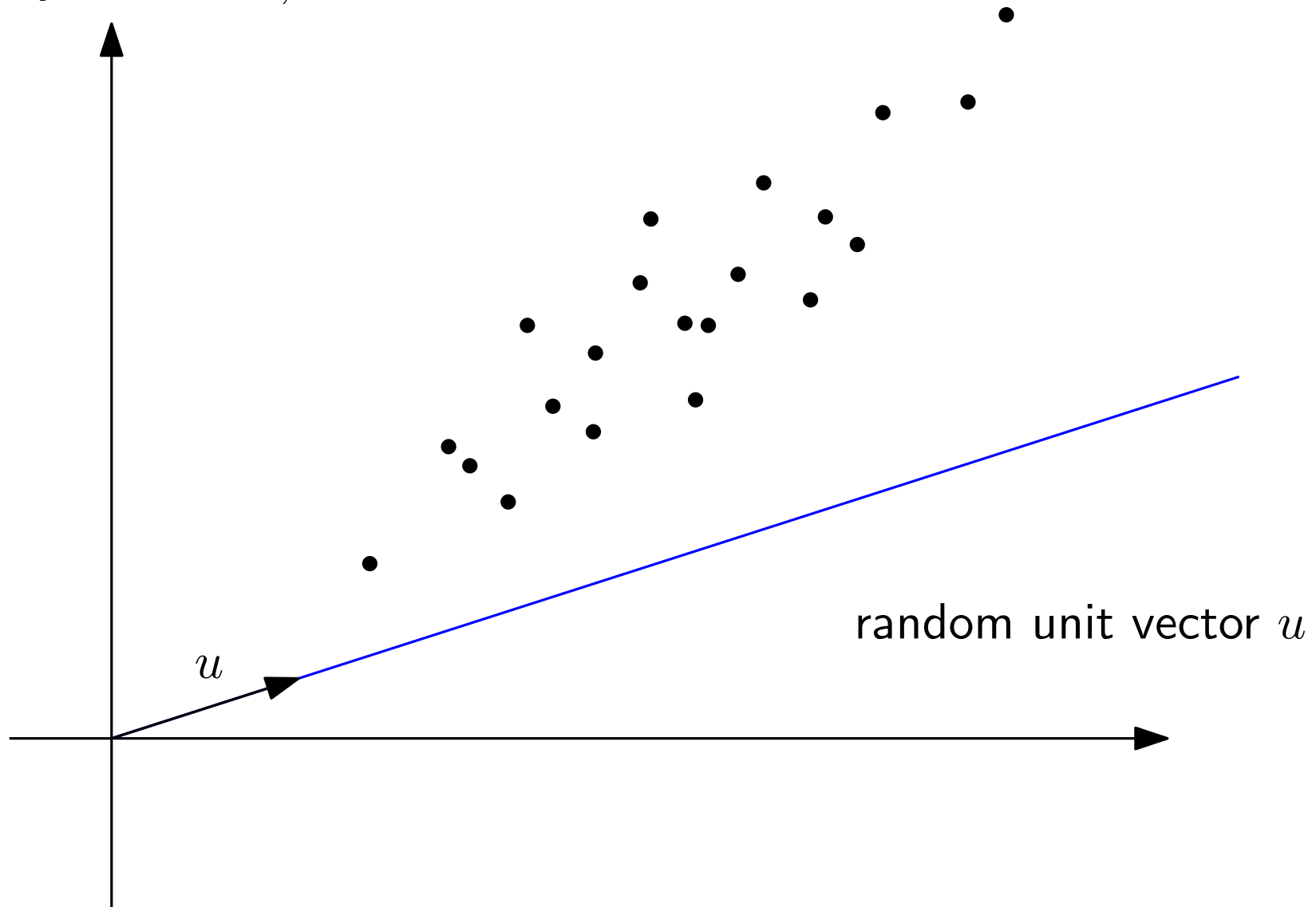
Principal Component Analysis (PCA)

Example: $d = 2, k = 1$



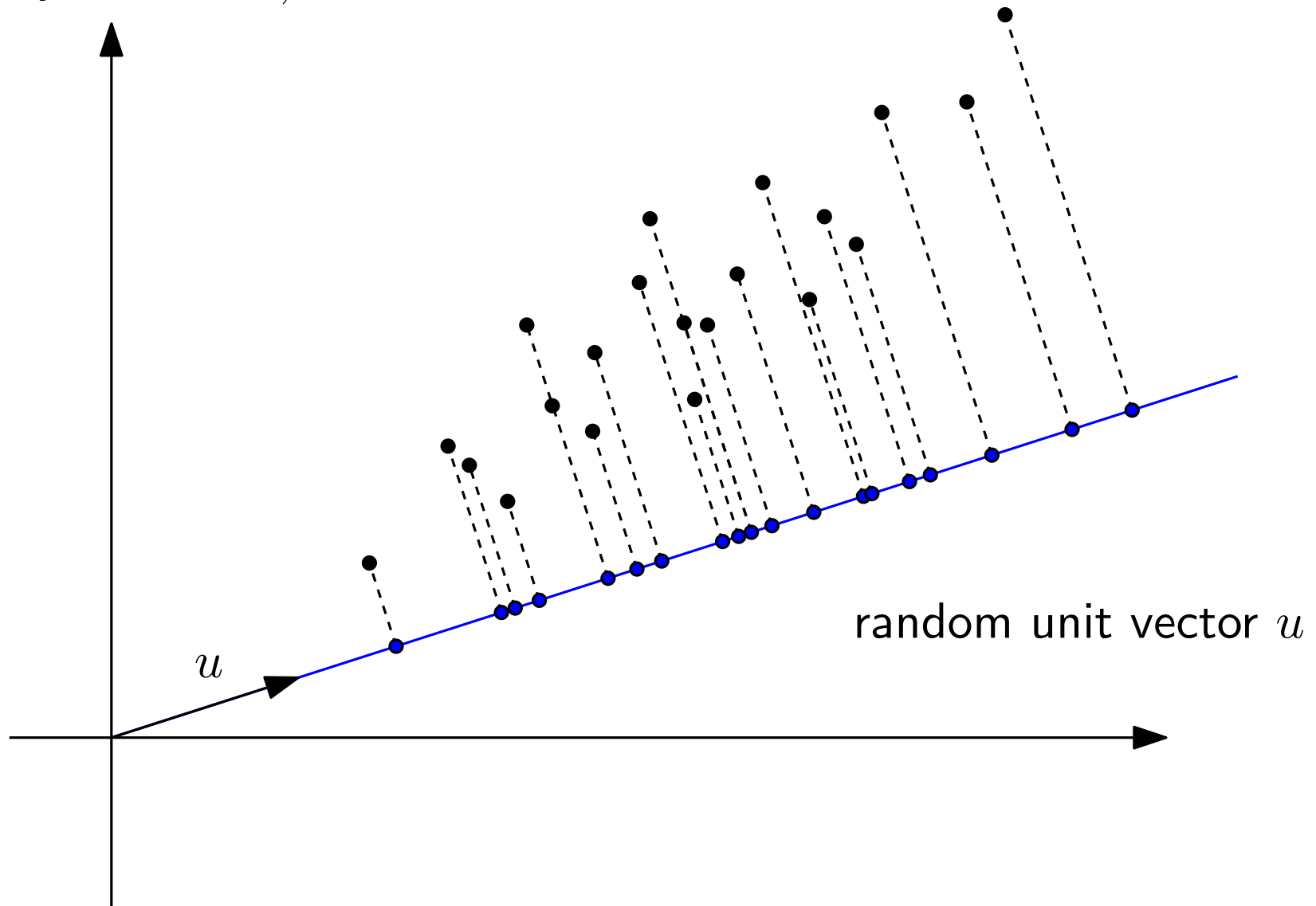
Random Projection

Example: $d = 2, k = 1$



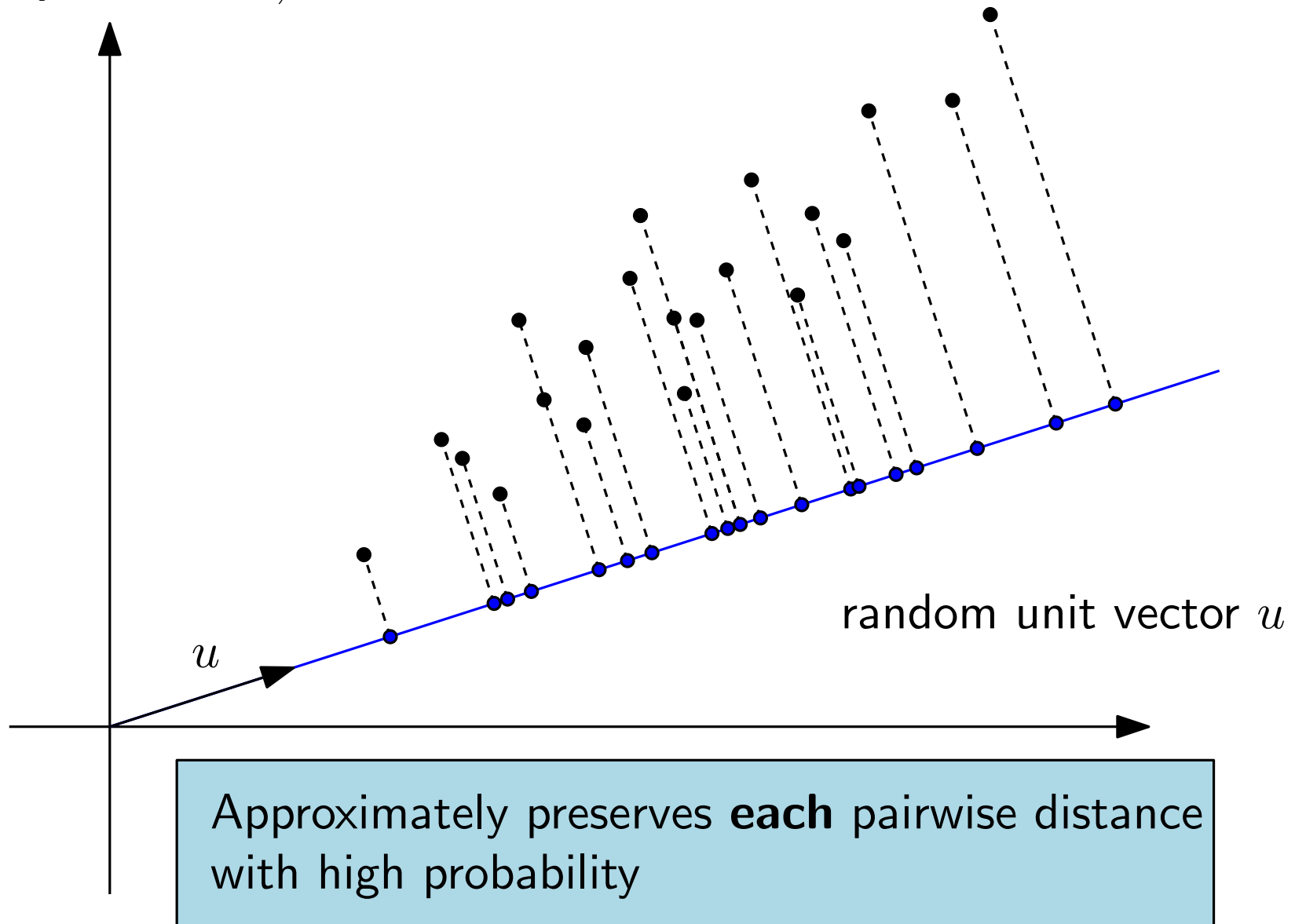
Random Projection

Example: $d = 2, k = 1$



Random Projection

Example: $d = 2, k = 1$



Embedding and Distortion

Given a point set $X \in \mathbb{R}^d$, we call a function $f : X \rightarrow \mathbb{R}^k$ an **embedding** of X . We define

$$\text{expansion}(f) = \max_{x, y \in X} \frac{\|f(x) - f(y)\|}{\|x - y\|}$$

$$\text{contraction}(f) = \max_{x, y \in X} \frac{\|x - y\|}{\|f(x) - f(y)\|}$$

The **distortion** of f is defined as the product of the expansion and the contraction of f .

Embedding and Distortion

Given a point set $X \in \mathbb{R}^d$, we call a function $f : X \rightarrow \mathbb{R}^k$ an **embedding** of X . We define

$$\text{expansion}(f) = \max_{x, y \in X} \frac{\|f(x) - f(y)\|}{\|x - y\|}$$

$$\text{contraction}(f) = \max_{x, y \in X} \frac{\|x - y\|}{\|f(x) - f(y)\|}$$

The **distortion** of f is defined as the product of the expansion and the contraction of f .

Note that for all $x, y \in X$ we have

$$\frac{1}{\beta} \|x - y\| \leq \|f(x) - f(y)\| \leq \alpha \|x - y\|$$

where α denotes the expansion and β denotes the contraction

Achlioptas' Random Projection (Algorithm)

Input: set of points $P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$, value of k

Output: set of points $Q = \{q_1, \dots, q_n\} \subseteq \mathbb{R}^k$

Algorithm:

- Generate a random $k \times d$ matrix \mathbf{R} by choosing

$$r_{i,j} = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

- For each $i = 1, \dots, n$, compute $q_i = \frac{1}{\sqrt{k}} \mathbf{R} p_i$

Achlioptas' Random Projection (Algorithm)

Input: set of points $P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$, value of k

Output: set of points $Q = \{q_1, \dots, q_n\} \subseteq \mathbb{R}^k$

Algorithm:

- Generate a random $k \times d$ matrix \mathbf{R} by choosing

$$r_{i,j} = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

- For each $i = 1, \dots, n$, compute $q_i = \frac{1}{\sqrt{k}} \mathbf{R} p_i$

Theorem:

Let $k_0 = \frac{4+2\beta}{\varepsilon^2/2-\varepsilon^3/3} \log n$, for given $\varepsilon, \beta > 0$. If $k \geq k_0$ then with probability at least $1 - \frac{1}{n^\beta}$, we have for all $p_i, p_j \in P$ that

$$(1 - \varepsilon) \|p_i - p_j\|^2 \leq \|q_i - q_j\|^2 \leq (1 + \varepsilon) \|p_i - p_j\|^2$$

History: Embedding Lemma

Random projections were invented by Johnson and Lindenstrauss.

Lemma (Johnson and Lindenstrauss, 1984):

Given $\varepsilon > 0$ and an integer n , let k be a positive integer $k \geq k_0 = O\left(\frac{\log n}{\varepsilon^2}\right)$. For every set of points

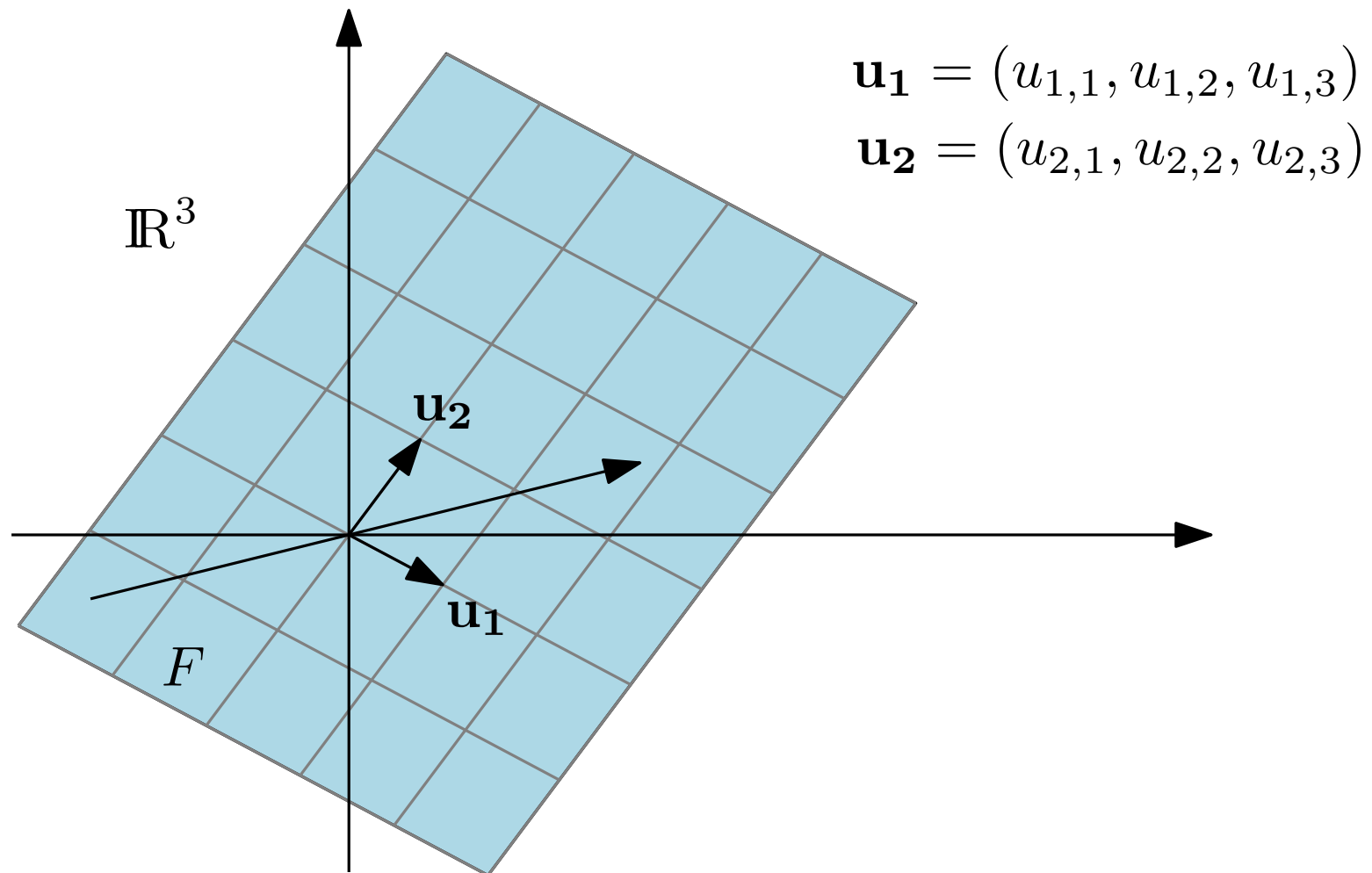
$P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$ there exists $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that for all $p_i, p_j \in P$

$$(1 - \varepsilon)\|p_i - p_j\|^2 \leq \|f(p_i) - f(p_j)\|^2 \leq (1 + \varepsilon)\|p_i - p_j\|^2.$$

Note: The proof uses a random projection to show that f exists. For historical reasons, the JL-lemma only talks about the existence of f .

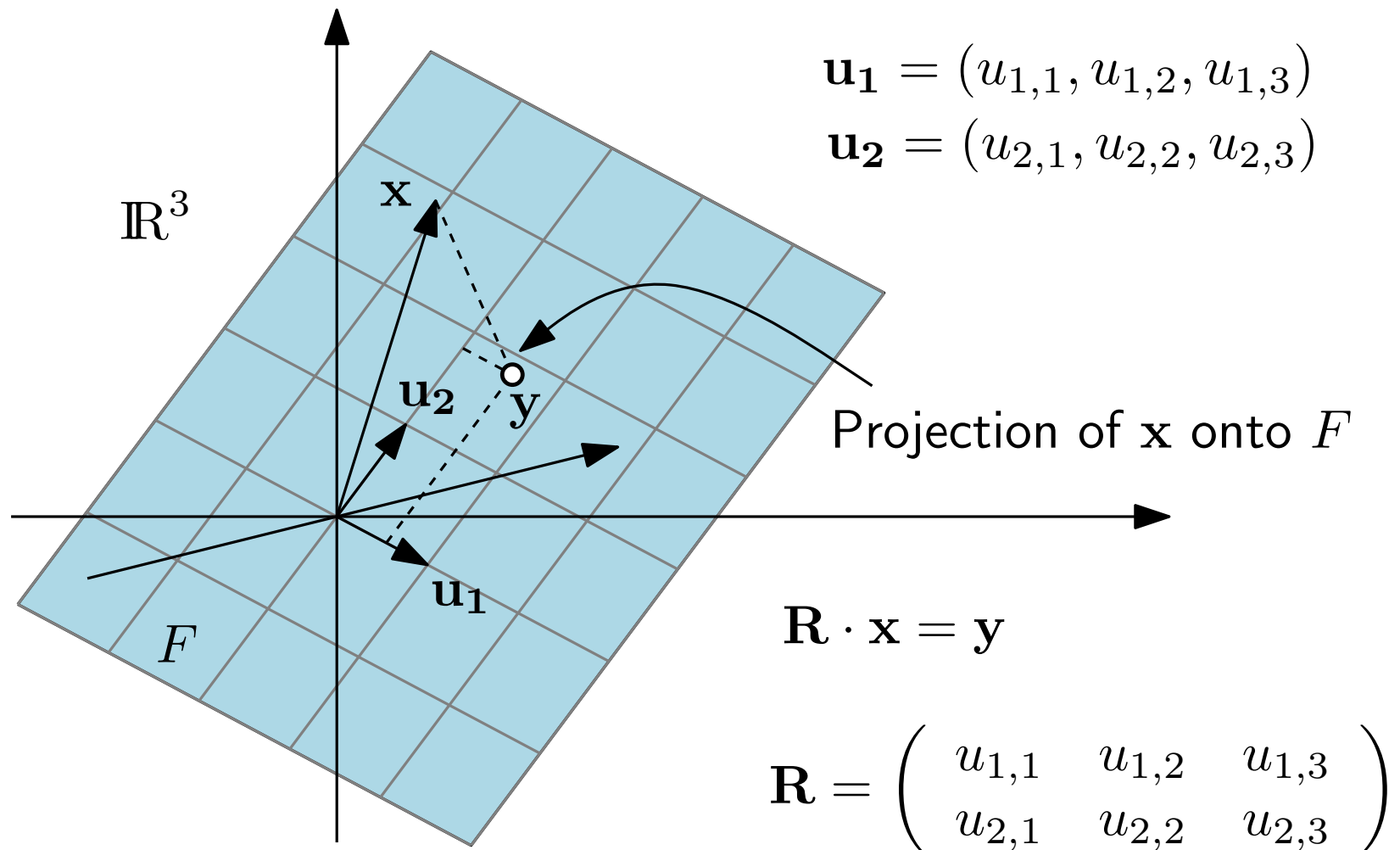
Linear Algebra: Projection onto subspace

Let F be a k -dimensional linear subspace of \mathbb{R}^d spanned by orthonormal vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ and let \mathbf{R} be the projection onto F



Linear Algebra: Projection onto subspace

Let F be a k -dimensional linear subspace of \mathbb{R}^d spanned by orthonormal vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ and let \mathbf{R} be the projection onto F



Linear Algebra: Projection onto subspace

A projection onto a subspace can be viewed as rotation followed by an axis-orthogonal projection

To see this, let's rewrite $\mathbf{R} = \mathbf{P} \cdot \mathbf{M}$ with

- \mathbf{M} : rotation to align each \mathbf{u}_i with standard basis vector \mathbf{v}_i
- \mathbf{P} : orthogonal projection onto first k coordinates

To find the rotation matrix \mathbf{M} , note that for $i = 1, \dots, k$

$$\mathbf{M} \cdot \mathbf{u}_i = \mathbf{v}_i \quad \Leftrightarrow \quad \mathbf{M}^{-1} \cdot \mathbf{v}_i = \mathbf{u}_i \quad \Leftrightarrow \quad \mathbf{M}^T \cdot \mathbf{v}_i = \mathbf{u}_i$$

- since \mathbf{v}_i is the i 'th standard basis vector, $\mathbf{M}^T \mathbf{v}_i$ is the i th column vector of \mathbf{M}^T
- thus, \mathbf{u}_i is the i th row vector of \mathbf{M} for $i = 1, \dots, k$

Analysis of the Random Rotation

We can think of Achlioptas transformation as a rotation \mathbf{M} followed by a projection \mathbf{P} onto the first k dimensions.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P}\mathbf{M}p$$

Example: $k = 2, d = 4$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{M} = \frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix}$$

Analysis of the Random Rotation

We can think of Achlioptas transformation as a rotation \mathbf{M} followed by a projection \mathbf{P} onto the first k dimensions.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P}\mathbf{M}p$$

Example: $k = 2, d = 4$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{M} = \frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix} \leftarrow \mathbf{R}$$

Analysis of the Random Rotation

We can think of Achlioptas transformation as a rotation \mathbf{M} followed by a projection \mathbf{P} onto the first k dimensions.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} p$$

Example: $k = 2, d = 4$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{M} = \frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix}$$

But is \mathbf{M} a rotation?

Analysis of the Random Rotation

\mathbf{M} is a rotation if and only if the product of \mathbf{M} and its transpose is the identity (i.e., \mathbf{M} is orthogonal)

$$\mathbf{M}\mathbf{M}^T = \mathbf{I}$$
$$\frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix} \frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{2,1} & r_{3,1} & r_{4,1} \\ r_{1,2} & r_{2,2} & r_{3,2} & r_{4,2} \\ r_{1,3} & r_{2,3} & r_{3,3} & r_{4,3} \\ r_{1,4} & r_{2,4} & r_{3,4} & r_{4,4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- (1) each pair of row vectors is orthogonal
- (2) each row vector has unit length

Let $\mathbf{r}_i = \frac{1}{\sqrt{d}}(r_{i,1}, \dots, r_{i,d})$ be the i th row vector of \mathbf{M}

Analysis of the Random Rotation

\mathbf{M} is a rotation if and only if the product of \mathbf{M} and its transpose is the identity (i.e., \mathbf{M} is orthogonal)

$$\mathbf{M}\mathbf{M}^T = \mathbf{I}$$

$$\frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{2,1} & r_{3,1} & r_{4,1} \\ r_{1,2} & r_{2,2} & r_{3,2} & r_{4,2} \\ r_{1,3} & r_{2,3} & r_{3,3} & r_{4,3} \\ r_{1,4} & r_{2,4} & r_{3,4} & r_{4,4} \end{pmatrix}$$

$$\frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- (1) each pair of row vectors is orthogonal
- (2) each row vector has unit length

Let $\mathbf{r}_i = \frac{1}{\sqrt{d}}(r_{i,1}, \dots, r_{i,d})$ be the i th row vector of \mathbf{M}

Analysis of the Random Rotation

\mathbf{M} is a rotation if and only if the product of \mathbf{M} and its transpose is the identity (i.e., \mathbf{M} is orthogonal)

$$\mathbf{M}\mathbf{M}^T = \mathbf{I}$$

$$\frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & r_{4,4} \end{pmatrix} \frac{1}{\sqrt{d}} \begin{pmatrix} r_{1,1} & r_{2,1} & r_{3,1} & r_{4,1} \\ r_{1,2} & r_{2,2} & r_{3,2} & r_{4,2} \\ r_{1,3} & r_{2,3} & r_{3,3} & r_{4,3} \\ r_{1,4} & r_{2,4} & r_{3,4} & r_{4,4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(1) each pair of row vectors is orthogonal

(2) each row vector has unit length

Let $\mathbf{r}_i = \frac{1}{\sqrt{d}}(r_{i,1}, \dots, r_{i,d})$ be the i th row vector of \mathbf{M}

Analysis of the Random Rotation

(continued)

Recall that each $r_{i,j}$ is a discrete random variable:

$$r_{i,j} = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

Let's analyze expected values of the matrix entries of $\mathbf{M}\mathbf{M}^T$

We will need the expected value of $r_{i,j}$:

$$\forall i, j : \mathbb{E}[r_{i,j}] = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0$$

Analysis of the Random Rotation

(continued)

(1) the expected angle between each pair of row vectors is orthogonal: $\forall i \neq j : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = 0$

Analysis of the Random Rotation

(continued)

(1) the expected angle between each pair of row vectors is orthogonal: $\forall i \neq j : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = 0$

Proof:

By linearity of expectation

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = \mathbb{E} \left[\sum_{t=1}^d \frac{r_{i,t}}{\sqrt{d}} \cdot \frac{r_{j,t}}{\sqrt{d}} \right] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t} \cdot r_{j,t}]$$

Analysis of the Random Rotation

(continued)

(1) the expected angle between each pair of row vectors is orthogonal: $\forall i \neq j : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = 0$

Proof:

By linearity of expectation

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = \mathbb{E} \left[\sum_{t=1}^d \frac{r_{i,t}}{\sqrt{d}} \cdot \frac{r_{j,t}}{\sqrt{d}} \right] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t} \cdot r_{j,t}]$$

for $i \neq j$ it holds that $r_{i,t}$ and $r_{j,t}$ are independent random variables, therefore

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_j \rangle] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t}] \cdot \mathbb{E} [r_{j,t}] = 0$$

□

Analysis of the Random Rotation

(continued)

(2) the expected squared length of each row vector is 1:

$$\forall i : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = 1$$

Analysis of the Random Rotation

(continued)

(2) the expected squared length of each row vector is 1:

$$\forall i : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = 1$$

Proof:

By linearity of expectation

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = \mathbb{E} \left[\sum_{t=1}^d \frac{r_{i,t}}{\sqrt{d}} \cdot \frac{r_{i,t}}{\sqrt{d}} \right] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t}^2]$$

Analysis of the Random Rotation

(continued)

(2) the expected squared length of each row vector is 1:

$$\forall i : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = 1$$

Proof:

By linearity of expectation

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = \mathbb{E} \left[\sum_{t=1}^d \frac{r_{i,t}}{\sqrt{d}} \cdot \frac{r_{i,t}}{\sqrt{d}} \right] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t}^2]$$

Note that $r_{i,t}^2$ is also a random variable and its expected value is:

$$\mathbb{E} [r_{i,t}^2] = (-1)^2 \cdot \frac{1}{2} + (1)^2 \cdot \frac{1}{2} = 1$$

Analysis of the Random Rotation

(continued)

(2) the expected squared length of each row vector is 1:

$$\forall i : \mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = 1$$

Proof:

By linearity of expectation

$$\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = \mathbb{E} \left[\sum_{t=1}^d \frac{r_{i,t}}{\sqrt{d}} \cdot \frac{r_{i,t}}{\sqrt{d}} \right] = \frac{1}{d} \sum_{t=1}^d \mathbb{E} [r_{i,t}^2]$$

Note that $r_{i,t}^2$ is also a random variable and its expected value is:

$$\mathbb{E} [r_{i,t}^2] = (-1)^2 \cdot \frac{1}{2} + (1)^2 \cdot \frac{1}{2} = 1$$

Therefore, $\mathbb{E} [\langle \mathbf{r}_i, \mathbf{r}_i \rangle] = 1$



Analysis of the Random Rotation

- We proved that \mathbf{M} is close to being a rotation on average.

Analysis of the Random Rotation

- We proved that \mathbf{M} is close to being a rotation on average.
- Therefore, we can think of f as a random rotation followed by an ordinary projection onto the first k coordinates.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} p$$

Analysis of the Random Rotation

- We proved that \mathbf{M} is close to being a rotation on average.
- Therefore, we can think of f as a random rotation followed by an ordinary projection onto the first k coordinates.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} p$$

- Next we want to analyze the effect of the random projection on a distance between two points

Analysis of the Random Rotation

- We proved that \mathbf{M} is close to being a rotation on average.
- Therefore, we can think of f as a random rotation followed by an ordinary projection onto the first k coordinates.

$$f(p) = \frac{1}{\sqrt{k}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \frac{1}{\sqrt{d}} \mathbf{R}p = \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P}\mathbf{M}p$$

- Next we want to analyze the effect of the random projection on a distance between two points
- Therefore we analyze for fixed $p_i, p_j \in P$ the expectation of its squared length in the projection

$$\mathbb{E} [\|f(p_i) - f(p_j)\|^2]$$

Analysis of a fixed distance d_{ij} (Expectation)

Claim: For fixed $p_i, p_j \in P$: $\mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \|p_i - p_j\|^2$

Analysis of a fixed distance d_{ij} (Expectation)

Claim: For fixed $p_i, p_j \in P$: $\mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \|p_i - p_j\|^2$

Proof:

Since f is a linear transformation, we have that

$$\mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \mathbb{E} [\|f(p_i - p_j)\|^2] = \mathbb{E} [\|f(\alpha)\|^2]$$

where $\alpha = (a_1, \dots, a_d) = p_i - p_j$.

Analysis of a fixed distance d_{ij} (Expectation)

Claim: For fixed $p_i, p_j \in P : \mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \|p_i - p_j\|^2$

Proof:

Since f is a linear transformation, we have that

$$\mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \mathbb{E} [\|f(p_i - p_j)\|^2] = \mathbb{E} [\|f(\alpha)\|^2]$$

where $\alpha = (a_1, \dots, a_d) = p_i - p_j$.

By the definition of f

$$\|f(\alpha)\|^2 = \left\| \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} \alpha \right\|^2 = \frac{d}{k} \|\mathbf{P} \mathbf{M} \alpha\|^2 = \frac{d}{k} \sum_{i=1}^k (\mathbf{r}_i \alpha)^2$$

where $\mathbf{r}_i = \frac{1}{\sqrt{d}}(r_{i,1}, \dots, r_{i,d})$ is the i th row vector of \mathbf{M} , as defined earlier.

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

by linearity of expectation

$$\mathbb{E} [\|f(\alpha)\|^2] = \mathbb{E} \left[\frac{d}{k} \sum_{i=1}^k (\mathbf{r}_i \alpha)^2 \right] = \frac{d}{k} \sum_{i=1}^k \mathbb{E} [(\mathbf{r}_i \alpha)^2]$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

by linearity of expectation

$$\mathbb{E} [\|f(\alpha)\|^2] = \mathbb{E} \left[\frac{d}{k} \sum_{i=1}^k (\mathbf{r}_i \alpha)^2 \right] = \frac{d}{k} \sum_{i=1}^k \mathbb{E} [(\mathbf{r}_i \alpha)^2]$$

by the definition of \mathbf{r}_i and α

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\left(\sum_{j=1}^d \frac{1}{\sqrt{d}} r_{i,j} a_j \right)^2 \right] = \mathbb{E} \left[\frac{1}{d} \left(\sum_{j=1}^d r_{i,j} a_j \right)^2 \right]$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

by linearity of expectation

$$\mathbb{E} [\|f(\alpha)\|^2] = \mathbb{E} \left[\frac{d}{k} \sum_{i=1}^k (\mathbf{r}_i \alpha)^2 \right] = \frac{d}{k} \sum_{i=1}^k \mathbb{E} [(\mathbf{r}_i \alpha)^2]$$

by the definition of \mathbf{r}_i and α

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\left(\sum_{j=1}^d \frac{1}{\sqrt{d}} r_{i,j} a_j \right)^2 \right] = \mathbb{E} \left[\frac{1}{d} \left(\sum_{j=1}^d r_{i,j} a_j \right)^2 \right]$$

We recursively expand the inner quadratic expression

$$\begin{aligned} \left(\sum_{j=1}^d r_{i,j} a_j \right)^2 &= (r_{i,1} a_1)^2 + 2r_{i,1} a_1 \left(\sum_{j=2}^d r_{i,j} a_j \right) + \left(\sum_{j=2}^d r_{i,j} a_j \right)^2 \\ &= \dots \\ &= \sum_{j=1}^d (r_{i,j} a_j)^2 + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2r_{i,j} a_j r_{i,l} a_l \end{aligned}$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d (r_{i,j} a_j)^2 + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2r_{i,j} a_j r_{i,l} a_l \right]$$

Note that $\alpha = (a_1, \dots, a_d)$ is a fixed vector, so it is not affected by randomness

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d (r_{i,j} a_j)^2 + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2r_{i,j} a_j r_{i,l} a_l \right]$$

Note that $\alpha = (a_1, \dots, a_d)$ is a fixed vector, so it is not affected by randomness

Therefore, by linearity of expectation

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \frac{1}{d} \sum_{j=1}^d a_j^2 \mathbb{E} [r_{i,j}^2] + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2a_j a_l \mathbb{E} [r_{i,j} r_{i,l}]$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d (r_{i,j} a_j)^2 + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2r_{i,j} a_j r_{i,l} a_l \right]$$

Note that $\alpha = (a_1, \dots, a_d)$ is a fixed vector, so it is not affected by randomness

Therefore, by linearity of expectation

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \frac{1}{d} \sum_{j=1}^d a_j^2 \underbrace{\mathbb{E} [r_{i,j}^2]}_{1 \text{ (as before)}} + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2a_j a_l \underbrace{\mathbb{E} [r_{i,j} r_{i,l}]}_{0 \text{ (as before)}}$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [(\mathbf{r}_i \alpha)^2] = \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d (r_{i,j} a_j)^2 + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2r_{i,j} a_j r_{i,l} a_l \right]$$

Note that $\alpha = (a_1, \dots, a_d)$ is a fixed vector, so it is not affected by randomness

Therefore, by linearity of expectation

$$\begin{aligned} \mathbb{E} [(\mathbf{r}_i \alpha)^2] &= \frac{1}{d} \sum_{j=1}^d a_j^2 \underbrace{\mathbb{E} [r_{i,j}^2]}_{1 \text{ (as before)}} + \sum_{j=1}^{d-1} \sum_{l=j+1}^d 2a_j a_l \underbrace{\mathbb{E} [r_{i,j} r_{i,l}]}_{0 \text{ (as before)}} \\ &= \frac{1}{d} \sum_{j=1}^d a_j^2 = \frac{1}{d} \|\alpha\|^2 \end{aligned}$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [\|f(\alpha)\|^2] = \frac{d}{k} \sum_{i=1}^k \mathbb{E} [(\mathbf{r}_i \alpha)^2] = \frac{d}{k} \sum_{i=1}^k \frac{1}{d} \|\alpha\|^2 = \|\alpha\|^2$$

Analysis of a fixed distance d_{ij} (Expectation)

(continued)

Plugging back into the equation..

$$\mathbb{E} [\|f(\alpha)\|^2] = \frac{d}{k} \sum_{i=1}^k \mathbb{E} [(\mathbf{r}_i \alpha)^2] = \frac{d}{k} \sum_{i=1}^k \frac{1}{d} \|\alpha\|^2 = \|\alpha\|^2$$

Now, plugging back the definition of α

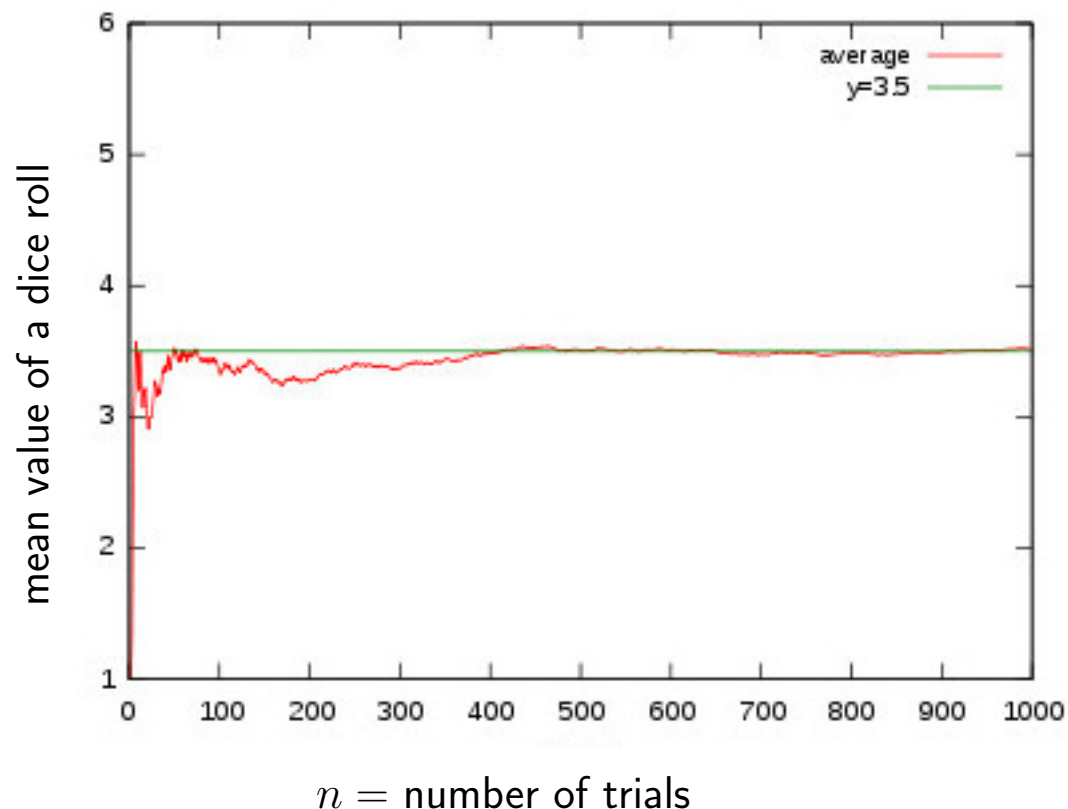
$$\mathbb{E} [\|f(p_i) - f(p_j)\|^2] = \|p_i - p_j\|^2$$



Law of Large Numbers

Let X_1, \dots, X_n be n samples of a random variable X .
The law of large numbers states that

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X] \right| \geq \varepsilon \right] \leq \frac{\text{Var}(X)}{n \cdot \varepsilon^2}$$



Concentration of mass in high dimensions

The unit hypercube: $[0, 1]^d = \{(x_1, \dots, x_d) \mid x_i \in [0, 1]\}$

Random vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$
(choose x_i independently and uniformly random in $[0, 1]$)

Consider the squared length $\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2$

Concentration of mass in high dimensions

The unit hypercube: $[0, 1]^d = \{(x_1, \dots, x_d) \mid x_i \in [0, 1]\}$

Random vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$
(choose x_i independently and uniformly random in $[0, 1]$)

Consider the squared length $\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2$

By the law of large numbers we have for high dimensions:

$$L := \frac{\|\mathbf{x}\|^2}{d} = \frac{1}{d} \sum_{i=1}^d x_i^2 \sim \frac{1}{3}$$

Concentration of mass in high dimensions

The unit hypercube: $[0, 1]^d = \{(x_1, \dots, x_d) \mid x_i \in [0, 1]\}$

Random vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$
(choose x_i independently and uniformly random in $[0, 1]$)

Consider the squared length $\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2$

By the law of large numbers we have for high dimensions:

$$L := \frac{\|\mathbf{x}\|^2}{d} = \frac{1}{d} \sum_{i=1}^d x_i^2 \sim \frac{1}{3}$$

This is the average of a random variable, since x_i are independent and identically distributed

(with $E[x_i^2] = \int_0^1 x_i^2 = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}$)

Concentration of mass in high dimensions

The unit hypercube: $[0, 1]^d = \{(x_1, \dots, x_d) \mid x_i \in [0, 1]\}$

Random vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$
(choose x_i independently and uniformly random in $[0, 1]$)

Consider the squared length $\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2 \sim \frac{d}{3}$
(for very large d)

By the law of large numbers we have for high dimensions:

$$L := \frac{\|\mathbf{x}\|^2}{d} = \frac{1}{d} \sum_{i=1}^d x_i^2 \sim \frac{1}{3}$$

↖ This is the average of a random variable, since x_i are independent and identically distributed

(with $E[x_i^2] = \int_0^1 x_i^2 = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}$)

Concentration of mass in high dimensions

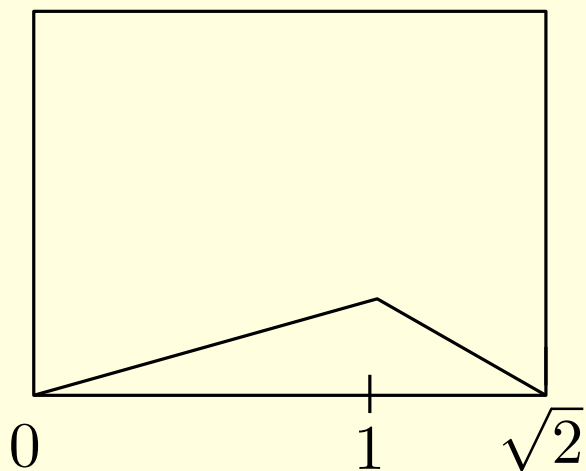
The unit hypercube: $[0, 1]^d = \{(x_1, \dots, x_d) \mid x_i \in [0, 1]\}$

Random vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$
(choose x_i independently and uniformly random in $[0, 1]$)

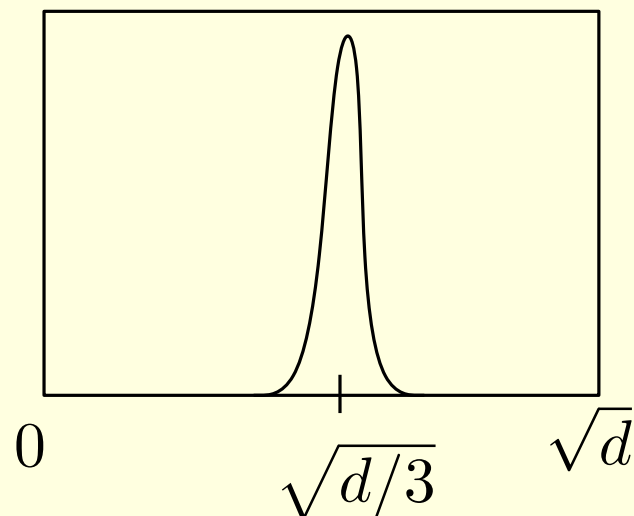
Consider the squared length $\|\mathbf{x}\|^2 = \sum_{i=1}^d x_i^2 \sim \frac{d}{3}$
(for very large d)

Distribution of the length of \mathbf{x} in low vs. high dimensions

$d = 2$



$d \gg 2$



Analysis of a fixed distance d_{ij}

Failure probability for two fixed points p_i, p_j :

$$\Pr \left[\frac{\|f(p_i) - f(p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

Analysis of a fixed distance d_{ij}

Failure probability for two fixed points p_i, p_j :

$$\Pr \left[\frac{\|f(p_i) - f(p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

by the linearity of f :

$$= \Pr \left[\frac{\|f(p_i - p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

$$= \Pr \left[\left\| \frac{f(p_i - p_j)}{\|p_i - p_j\|} \right\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

$$= \Pr \left[\left\| f \left(\frac{p_i - p_j}{\|p_i - p_j\|} \right) \right\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

Analysis of a fixed distance d_{ij}

Failure probability for two fixed points p_i, p_j :

$$\Pr \left[\frac{\|f(p_i) - f(p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

by the linearity of f :

$$= \Pr \left[\frac{\|f(p_i - p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

$$= \Pr \left[\left\| \frac{f(p_i - p_j)}{\|p_i - p_j\|} \right\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

$$= \Pr \left[\left\| f \left(\frac{p_i - p_j}{\|p_i - p_j\|} \right) \right\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

define $\alpha = \frac{p_i - p_j}{\|p_i - p_j\|}$ (note that α is a fixed unit vector)

$$= \Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

squared length of the **random projection** of a unit vector

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

squared length of the **random projection** of a unit vector

By the definition of f :

$$\|f(\alpha)\|^2 = \left\| \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} \alpha \right\|^2 = \frac{d}{k} \|\mathbf{P} \mathbf{M} \alpha\|^2$$

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

squared length of the **random projection** of a unit vector

By the definition of f :

$$\|f(\alpha)\|^2 = \left\| \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} \alpha \right\|^2 = \frac{d}{k} \|\mathbf{P} \mathbf{M} \alpha\|^2$$

For simplicity, assume \mathbf{M} is a proper random rotation.

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

squared length of the **random projection** of a unit vector

By the definition of f :

$$\|f(\alpha)\|^2 = \left\| \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} \alpha \right\|^2 = \frac{d}{k} \|\mathbf{P} \mathbf{M} \alpha\|^2$$

For simplicity, assume \mathbf{M} is a proper random rotation.

In this case, $\mathbf{M} \alpha$ is a **random unit vector**

Analysis of a fixed distance d_{ij}

(continued)

$$\Pr \left[\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon] \right]$$

squared length of the **random projection** of a unit vector

By the definition of f :

$$\|f(\alpha)\|^2 = \left\| \frac{\sqrt{d}}{\sqrt{k}} \mathbf{P} \mathbf{M} \alpha \right\|^2 = \frac{d}{k} \|\mathbf{P} \mathbf{M} \alpha\|^2$$

For simplicity, assume \mathbf{M} is a proper random rotation.

In this case, $\mathbf{M} \alpha$ is a **random unit vector**

squared length of a **random unit vector**, projected onto the first k coordinates

Analysis of a fixed distance d_{ij}

Using concentration of measure, Achlioptas shows the following lemma (we omit the full proof):

Lemma:

Let $r_{i,j}$ be chosen uniformly random from $\{-1, 1\}$, then for any $\varepsilon > 0$ and any unit vector $\alpha \in \mathbb{R}^d$,

$$\Pr [\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon]] < 2 \cdot e^{(-\frac{k}{2}(\varepsilon^2/2 - \varepsilon^3/3))}$$

Therefore, choosing

$$k \geq \frac{4 + 2\beta}{\varepsilon^2/2 + \varepsilon^3/3} \log n$$

is sufficient to ensure

$$\Pr \left[\frac{\|f(p_i) - f(p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right] = \Pr [\|f(\alpha)\|^2 \notin [1 - \varepsilon, 1 + \varepsilon]] < \frac{2}{n^{2+\beta}}$$

Analysis of the Distortion

Theorem:

Let $k_0 = \frac{4+2\beta}{\varepsilon^2/2-\varepsilon^3/3} \log n$, for given $\varepsilon, \beta > 0$. If $k \geq k_0$ then with probability at least $1 - \frac{1}{n^\beta}$, we have for all $p_i, p_j \in P$ that

$$(1 - \varepsilon)\|p_i - p_j\|^2 \leq \|f(p_i) - f(p_j)\|^2 \leq (1 + \varepsilon)\|p_i - p_j\|^2$$

Analysis of the Distortion

Theorem:

Let $k_0 = \frac{4+2\beta}{\varepsilon^2/2-\varepsilon^3/3} \log n$, for given $\varepsilon, \beta > 0$. If $k \geq k_0$ then with probability at least $1 - \frac{1}{n^\beta}$, we have for all $p_i, p_j \in P$ that

$$(1 - \varepsilon) \|p_i - p_j\|^2 \leq \|f(p_i) - f(p_j)\|^2 \leq (1 + \varepsilon) \|p_i - p_j\|^2$$

Proof :

- there are $\binom{n}{2}$ pairs of points in P
- bound the failure probability for two fixed points:

$$\Pr \left[\frac{\|f(p_i) - f(p_j)\|^2}{\|p_i - p_j\|^2} \notin [1 - \varepsilon, 1 + \varepsilon] \right] < \frac{2}{n^{2+\beta}}$$

- apply the union bound: $\binom{n}{2} \cdot \frac{2}{n^{2+\beta}} < \frac{n^2}{2} \cdot \frac{2}{n^{2+\beta}} = \frac{1}{n^\beta}$

Alternative Projection Matrix

Note:

For generating the matrix \mathbf{R} , other distributions are possible.
In particular,

Alternative Projection Matrix

Note:

For generating the matrix \mathbf{R} , other distributions are possible.
In particular,

$$r_{i,j} = \sqrt{3} \begin{cases} +1 & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -1 & \text{with probability } \frac{1}{6} \end{cases}$$

Alternative Projection Matrix

Note:

For generating the matrix \mathbf{R} , other distributions are possible.
In particular,

$$r_{i,j} = \sqrt{3} \begin{cases} +1 & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -1 & \text{with probability } \frac{1}{6} \end{cases}$$

What would be the advantage of this?

Alternative Projection Matrix

Note:

For generating the matrix \mathbf{R} , other distributions are possible.
In particular,

$$r_{i,j} = \sqrt{3} \begin{cases} +1 & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -1 & \text{with probability } \frac{1}{6} \end{cases}$$

What would be the advantage of this?

This generates a sparser matrix

Summary

- Nearest-neighbor searching
- Embedding and Distortion
- Achlioptas' Random Projection
- Projection onto a subspace
- Random Rotation (Expectation)
- Analysis of a fixed distance (Expectation)
- Law of large numbers
- Concentration of measure
- Analysis of a fixed distance
- Analysis of the Distortion
- Alternative Projection Matrix

References

- Dimitris Achlioptas: "Database-friendly random projections: Johnson-Lindenstrauss with binary coins" *Journal of Computer and System Sciences* 66 (2003) 671687
- Ashraf M. Kibriya and Eibe Frank: "An Empirical Comparison of Exact Nearest Neighbour Algorithms"
- Avrim Blum, John Hopcroft, Ravindran Khannan: *Foundations of Data Science*