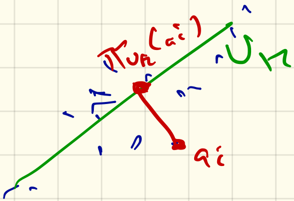
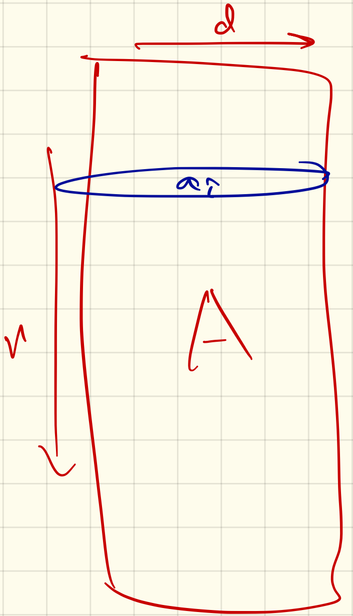



# Random Projections

Input $n$ data points $\{a_1, a_2, \dots, a_n\}$

$a_i \in \mathbb{R}^d$




SVD / PCA $\rightarrow A_k \in \mathbb{R}^{n \times d}$

↳ Best rank-$k$ subspace

$$A = \overset{k}{U}\, S\, V^T$$

top $k$ left sing. vectors

$$U^* = U_k = \underset{\mathrm{rank}(U) \le k}{\arg\min} \sum_{a_i \in A} \| a_i - \pi_{U_k}(a_i) \|^2$$

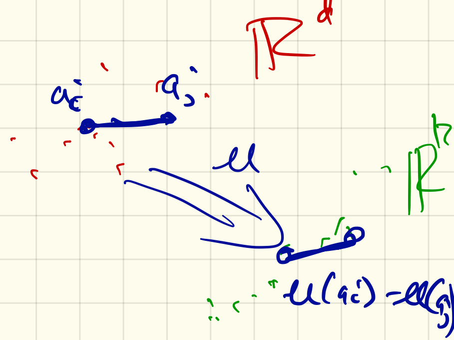
$U_k =$

$|A| = n$

Define some map $u : \mathbb{R}^d \to \mathbb{R}^k$

$\forall a_i, a_j \in A$ error parameter $\varepsilon > 0$
$\approx 0.10$

$$(1-\varepsilon)\|a_i - a_j\| \le \|u(a_i) - u(a_j)\| \le (1+\varepsilon)\|a_i - a_j\|$$




with prob. $\ge 1-\delta$

$$k = \approx \left(\frac{1}{\varepsilon^2}\right) \log \frac{n}{\delta}$$

How to construct $\mu : \mathbb{R}^d \to \mathbb{R}^k$

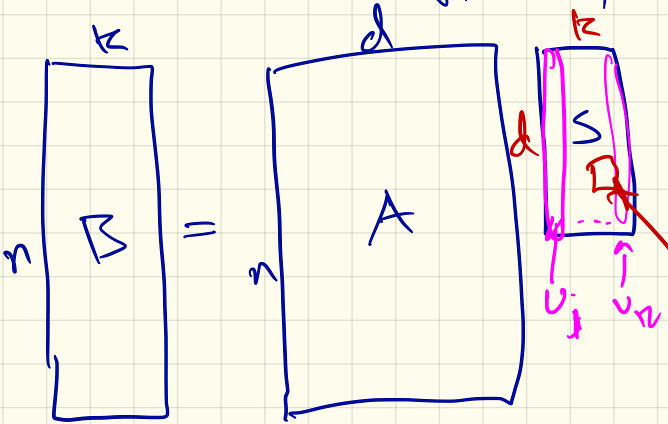
$$\mu(a) = (\mu_1(a), \mu_2(a), \dots \mu_k(a))$$

Focus on $\mu_j(a)$

$$\mu_j(a) = \frac{\sqrt{d}}{\sqrt{k}} \langle a, u_j \rangle$$

so $u_j \in \text{Unif}(S^{d-1})$

$\|u_j\| = 1$




$g_j \sim N_d(0, 1)$ ← Box-Muller

$$u_j = \frac{g_j}{\|g_j\|}$$

$S_{ij} \sim \frac{\sqrt{d}}{\sqrt{k}} N_1(0, 1)$

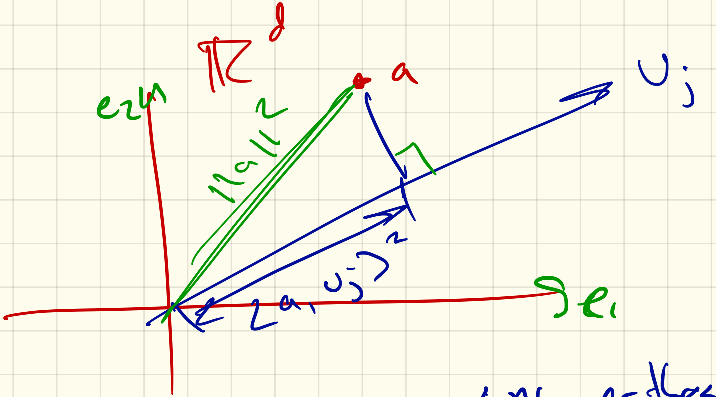
$$E\left[\|u(a)\|^2\right] = \|a\|^2$$

$$E\left[\left(\frac{\sqrt{d}}{\sqrt{k}}\langle u_j, a\rangle\right)^2\right]$$

$$E\left[d\langle u_j, a\rangle^2\right] = \|a\|^2$$

jth coord

$$e_j = (0, 0, \dots, 1, 0 \dots 0)$$



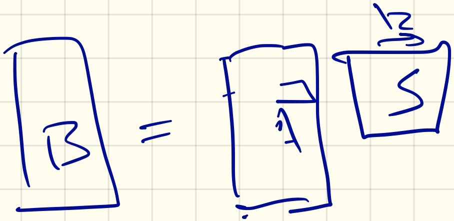

$$\|a\|^2 = \sum_{j=1}^{d} \langle a, e_j\rangle^2$$

Pythagorian

any orthon basis

$$U = [u_1, u_2 \dots u_j \dots u_d]$$

# Oblivious Subspace Embedding




$$k = O\left(\frac{d}{\varepsilon^2}\log\frac{1}{\delta}\right)$$

$\forall x \in \mathbb{R}^d$

$$(1-\varepsilon)\|Ax\| \leq \|Bx\| \leq (1+\varepsilon)\|Ax\|$$

## Count-Sketch