$\boxed{\text{21 Mar 2022}}$ SVD on i.i.d. random vectors (continued)

We have random vectors $a_1, \ldots, a_n \in \mathbb{R}^d$ sampled i.i.d. (independently and identically distributed) from a distribution with covariance matrix

$$\Sigma = BB^T.$$

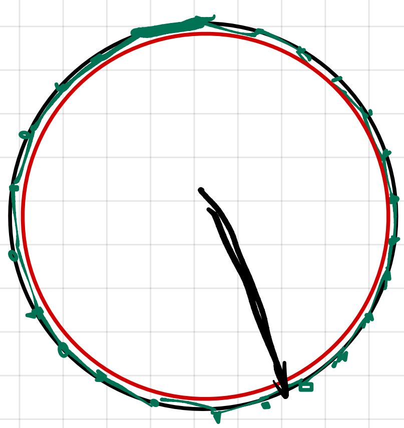
(On Friday we assumed Gaussian. In a different equally valid application, it's the unif distrib over columns of a matrix with d rows and $N \gg n$ columns.)

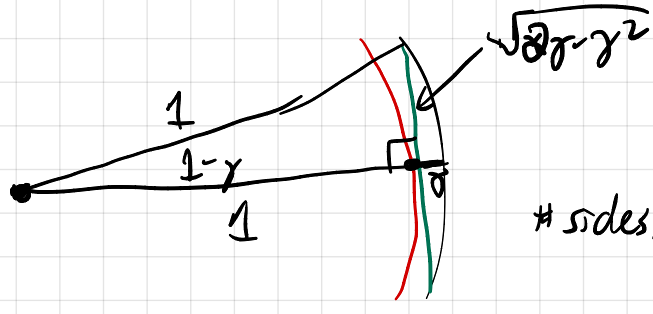
"downsampling" or "sparsifying" a dataset, from N vectors down to n.

Given $\varepsilon, \delta > 0$ we want to understand how large must n be so that top right singular vector $\hat{v}$ of A^T is close to top right sing vector of B^T, v_1, in the sense that $\langle \hat{v}, v_1 \rangle > 1-\varepsilon$ with prob. $1-\delta$.

$\hat{v}$ solves $\max \| A^T \hat{v} \|_2$ s.t. $\| \hat{v} \|_2 = 1$

v_1 solves $\max \| \mathbb{E}(A^T) v_1 \|_2$ s.t. $\| v_1 \|_2 = 1$





$\#\text{sides} \approx \dfrac{\pi}{\sqrt{2\gamma}}$

Constructing a set of points in $\mathbb{R}^d$ that lie on a radius 1 sphere and contain a radius $1-\gamma$ sphere in their convex hull.

$(\gamma \ll 1)$

A construction that works: $k \triangleq \left\lceil \frac{d}{2\gamma} \right\rceil$.

for set of indices $I = (i_1, \dots, i_d)$

such that $i_1, \dots, i_d \geq 0$

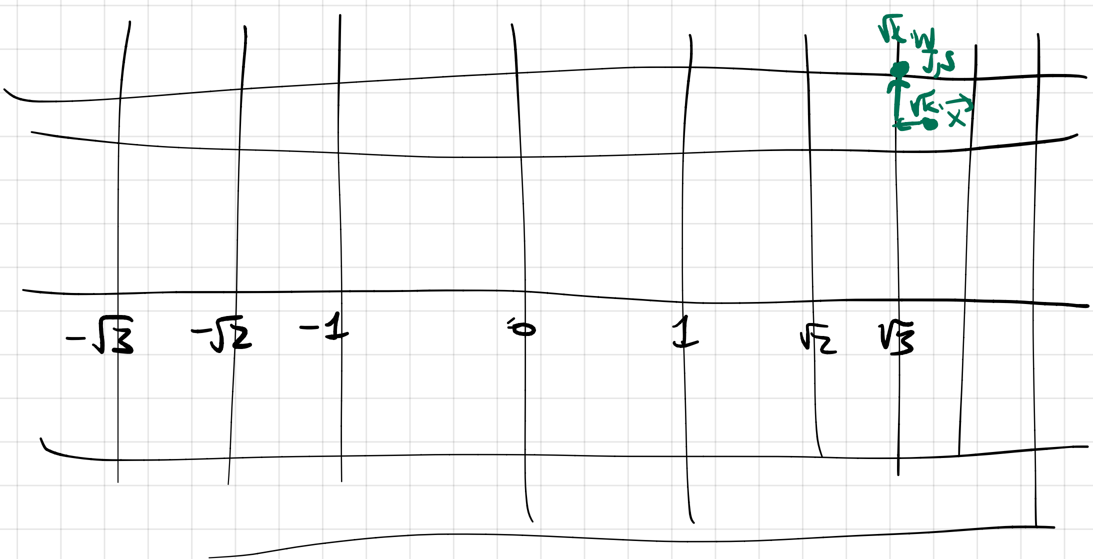
$i_1 + \dots + i_d = k$

and a set of signs $S = (s_1, \dots, s_d) \in \{\pm 1\}^d$

let

$$W_{I,S} = \frac{1}{\sqrt{k}} \begin{bmatrix} s_1 \sqrt{i_1} \\ s_2 \sqrt{i_2} \\ \vdots \\ s_d \sqrt{i_d} \end{bmatrix}.$$

$$\| W_{I,S} \|_2^2 = \frac{1}{k} \cdot \sum_{j=1}^{d} i_j = 1.$$



Lemma: If $\|x\|_2 = 1$, $\exists$ I, S s.t. $\langle w_{I,S}, x\rangle \geq 1-\gamma$.

Proof. For $j = 1, \ldots, d$ let $S_j = \begin{cases} +1 & \text{if } x_j \geq 0 \\ -1 & \text{if } x_j < 0. \end{cases}$

For $j = 1, \ldots, d-1$ let $i_j = \lfloor k x_j^2 \rfloor$.

Choose i_d s.t. $i_1 + \ldots + i_d = k$.

Do some calculations of rounding errors and conclude $\langle w_{I,S}, x\rangle > 1-\gamma$.

How many $w_{I,S}$ vectors are there?

sign patterns, S, is 2^d.

index vectors $\vec{i}$ $i_1 + \ldots + i_d = k$

$$\binom{d+k-1}{d-1} < \binom{2k}{d} < \left(\frac{2ek}{d}\right)^d$$

In total we have $\left(\frac{4ek}{d}\right)^d$ pairs I, S.

Recall $\frac{k}{d} = \frac{1}{2\gamma}$, so $\left(\frac{2e}{\gamma}\right)^d$.