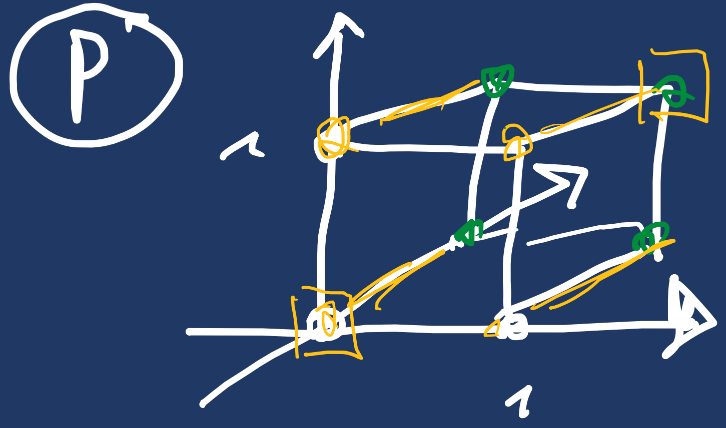
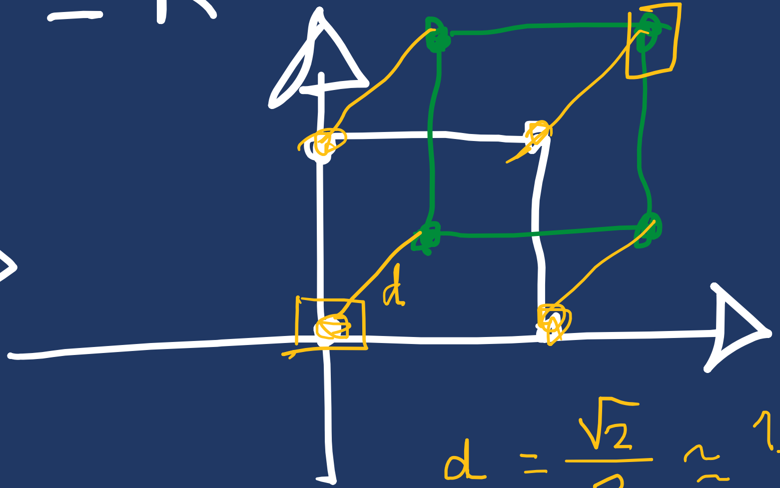
Johnson–Lindenstrauss theorem.



$d(\square, \square) = \sqrt{3}$

$C = \{0,1\}^3 \subseteq R^3$

$R^3 \xrightarrow{\pi}$



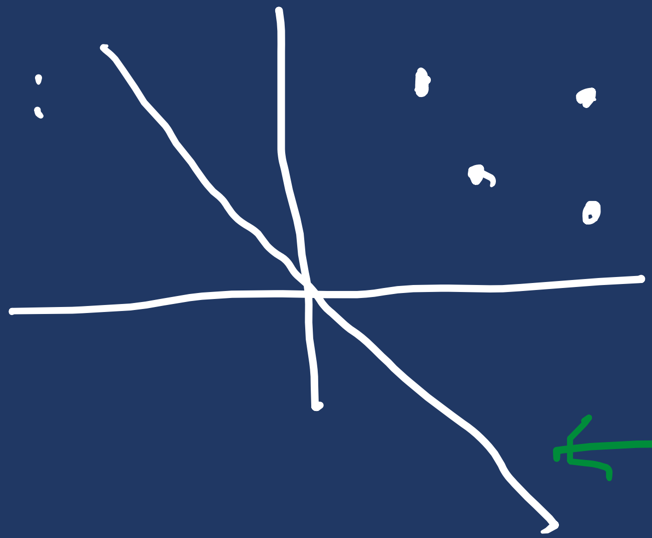
$d = \frac{\sqrt{2}}{2} \approx \frac{1.41}{2} \approx 0.7$

$$(1-\varepsilon)\,\|P-Q\|^2 \leq \|\pi(P)-\pi(Q)\|^2 \leq (1+\varepsilon)\,\|P-Q\|^2$$

Q : what is the distortion ε?

Q : can we do it better.

Idea:



$X \subseteq \mathbb{R}^d$

$|X| = n$

S - rand subspace of $\mathbb{R}^d$ of dim k

use π_S ← proj. on S.

$P, Q \in X$

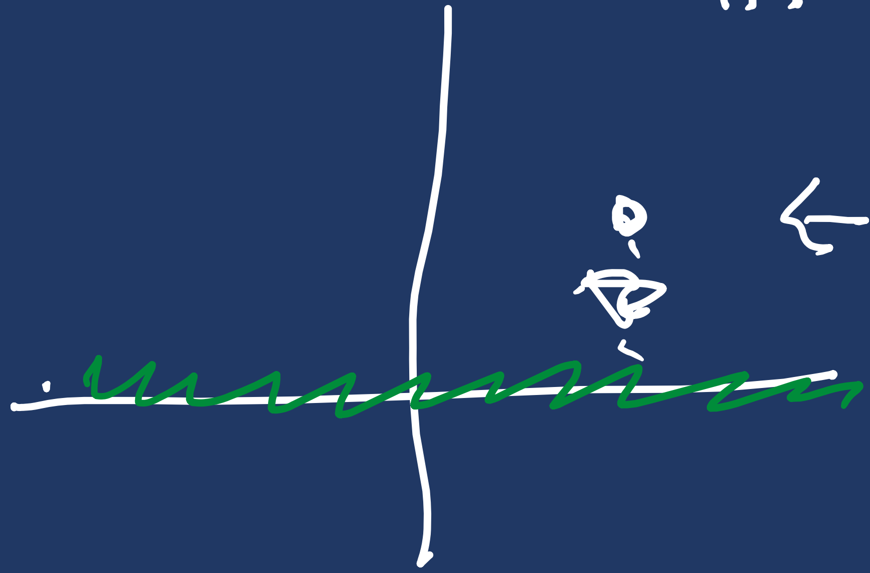
$$\|\pi_S(P) - \pi_S(Q)\| = \|\pi_S(\underbrace{P-Q}_{u})\|$$

$$= \pi_S(u)$$

• u — point in R^d

← S ← rand ~~plane~~ k-dim subspace

$|||$



← random point in R^d

fixed subspace k-dim.

$$S = \{(x_1, \dots, x_k, 0, 0, \dots, 0) : x_1 \dots x_k \in R\}$$

$$\pi_S(\alpha \cdot \vec{u}) = \alpha\, \pi_S(\vec{u})$$

we may think only about vectors of length 1.

$$u \rightsquigarrow \frac{u}{\|u\|} \in S^{d-1} = \{\bar{x} \in \mathbb{R}^d : \|\bar{x}\| = 1\}$$

How to generate a random point from S^{d-1}?

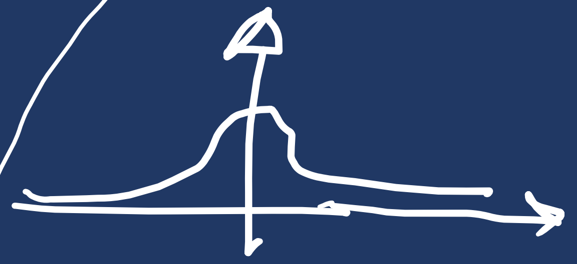
Proper solution

- take $x_1, \dots, x_d \sim \mathcal{N}(0,1)$ indep.
- form $X = [x_1, \dots, x_d]$
- consider $Y = \frac{X}{\|X\|}$.

$X \sim \mathcal{N}(0, 1)$: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

! $\int_{\mathbb{R}} e^{-x^2} dx = ?$

! how to calculate it.



Gaussian curve

$X_1, \dots, X_d \sim \mathcal{N}(0,1)$, ind.

$$f(x_1, \dots, x_d) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}} e^{-x_d^2/2}$$

$$= (2\pi)^{-d/2} e^{-(x_1^2 + \dots + x_d^2)/2}$$

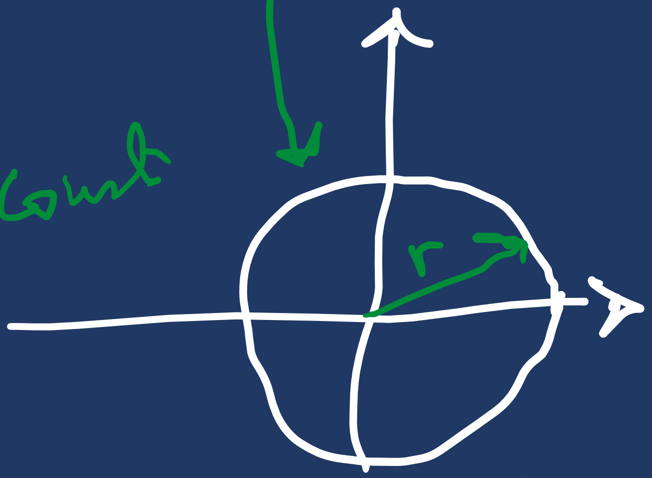
$$= (2\pi)^{-d/2} e^{-\|(x_1 \dots x_d)\|^2/2}$$

depends only on $\|(x_1 \dots x_d)\|$

inv. under rotations.

$d=2$

const



$f = (2\pi)^{-1} e^{-r^2/2}$

Suppose $X_1, \ldots, X_d \sim \mathcal{N}(0,1)$, indep.

$$Y = \frac{[X_1 \ldots X_d]}{\|[X_1, \ldots X_d]\|} = [y_1, y_2, \ldots, y_d]$$

$$1 = y_1^2 + \ldots + y_d^2$$

$$1 = E[1] = E[y_1^2 + \ldots + y_d^2] = \sum_{i=1}^{d} E\left[\frac{X_i^2}{X_1^2 + \ldots + X_d^2}\right],$$

↑ the same

$$E\left[\frac{X_i^2}{X_1^2 + \ldots + X_d^2}\right] = \frac{1}{d}$$

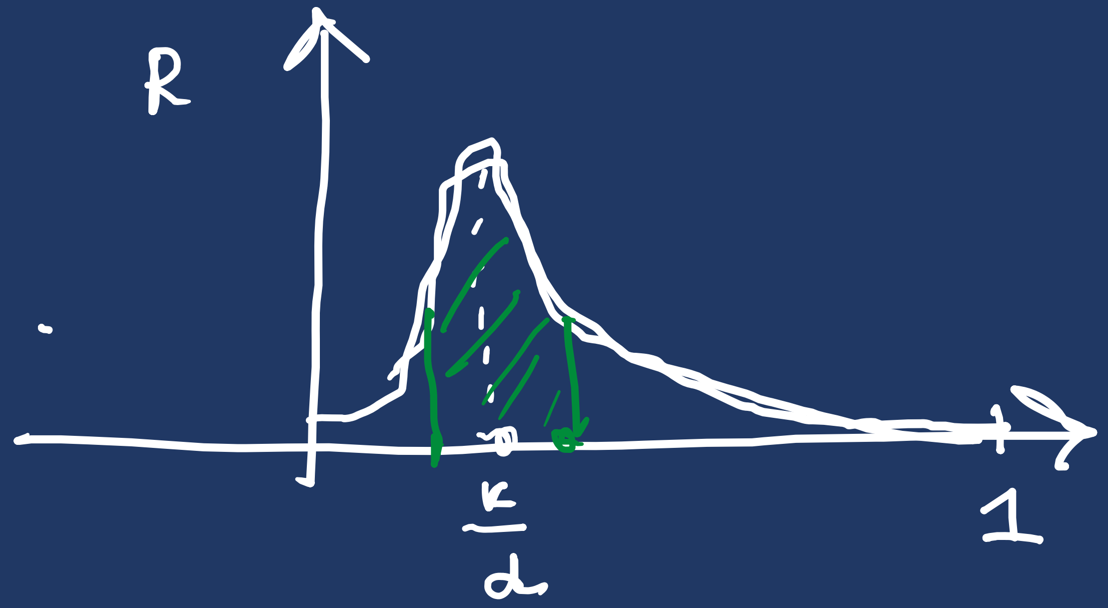
Take $Z = [y_1, \dots, y_k]$.

Put $L = \|Z\|^2$

$$E[L] = E\left[\sum_{i=1}^{k} \frac{x_i^2}{x_1^2 + \dots + x_d^2}\right] = \sum_{i=1}^{k} E\left[\frac{x_i^2}{x_1^2 + \dots + x_d^2}\right]$$

$$= \frac{k}{d}$$

$$E[L] = \frac{k}{d}$$



density of L

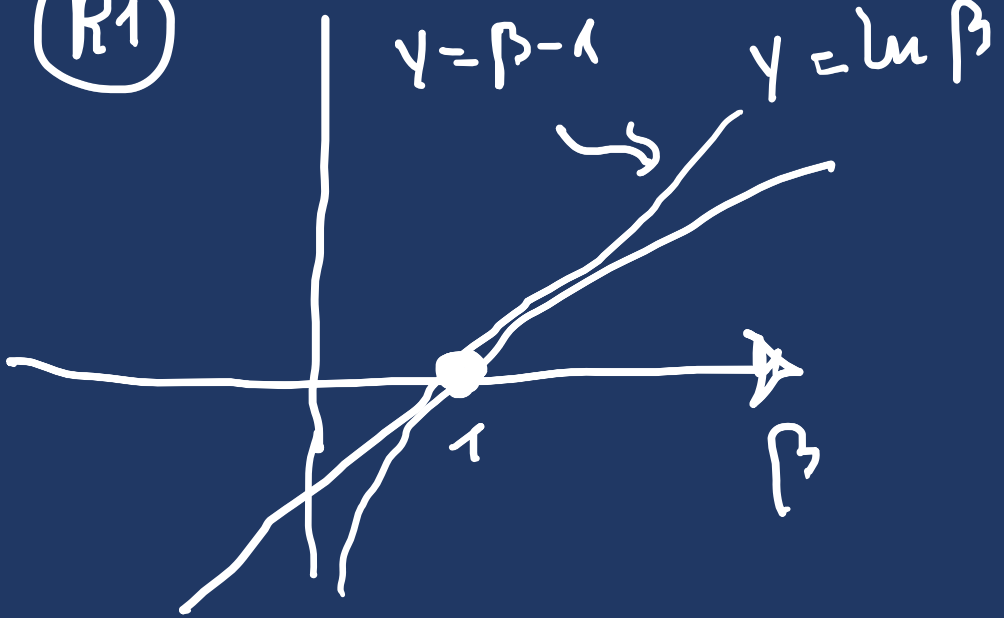
Lemma: If $0<\beta<1$ then

$$\Pr\left[L \le \beta \frac{k}{d}\right] \le \exp\left(\frac{k}{2}\overbrace{(1-\beta+\ln\beta)}^{\text{negative}}\right)$$

If $\beta>1$ then

$$\Pr\left[L \ge \beta \frac{k}{d}\right] \le \underbrace{\exp\left(\frac{k}{2}(1-\beta+\ln\beta)\right)}_{\text{!!! does not depend on } d \text{ !!}}$$

R1



$$(\ln\beta)'\Big|_{\beta=1} = \left(\frac{1}{\beta}\right)_{\beta=1} = 1$$

$$(\ln\beta)'' = -\frac{1}{\beta^2} < 0$$

$$\ln\beta \le \beta - 1$$

$$(1-\beta)+\ln\beta \le 0$$

"An Elementary proof of a Theorem of Johnson - Lind.."

S. Dasgupta, A. Gupta.

$$\ln \frac{1}{1-x} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots = \sum_{i=1}^{\infty} \frac{x^i}{i} \qquad |x| < 1$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

$$k \geq \frac{4}{\varepsilon^2/2 - \varepsilon^3/3} \ln n$$

Chernoff bounds

For $v \in \mathbb{R}^d$: $v' =$ orth. proj. of v onto a random subspace of dim k.

Take $v_1, v_2 \in X$; $u = v_1 - v_2$; $u' = (v_1 - v_2)' =$
$= v_1' - v_2'$.

$$L = \frac{\|u'\|^2}{\|u\|^2} \qquad E[L] = \frac{k}{d}$$

$$P\left[L \leq \beta \cdot \frac{k}{d}\right] = \Pr\left[\frac{\|u'\|^2}{\|u\|^2} \leq \beta \frac{k}{d}\right] =$$

$$= P\left[\|u'\|^2 \leq \beta \frac{k}{d} \cdot \|u\|^2\right] =$$

$f(u) = \sqrt{\frac{d}{k}} \cdot u'$

$$= P\left[\frac{d}{k} \cdot \|u'\|^2 \leq \beta \|u\|^2\right] =$$

$$= P\left[\left\|\sqrt{\frac{d}{k}}\, u'\right\|^2 \leq \beta \|u\|^2\right]$$

$$P\left[L \le \beta \frac{k}{d}\right] \le \exp\left(\frac{k}{2}\left(1-\beta+\ln\beta\right)\right) =$$

$\beta \leftarrow 1-\varepsilon$

$$= \exp\left(\frac{k}{2}\left(1-(1-\varepsilon)+\ln(1-\varepsilon)\right)\right) = \exp\left(\frac{k}{2}\left(\varepsilon+\ln(1-\varepsilon)\right)\right)$$

$$\ln(1-\varepsilon) \le -\varepsilon - \frac{\varepsilon^2}{2} \qquad (\varepsilon \in [0,1)) \quad \text{easy}$$

$$\le \exp\left(\frac{k}{2}\left(\varepsilon + \left(-\varepsilon - \varepsilon^2/2\right)\right)\right) = \exp\left(-\frac{k}{2}\frac{\varepsilon^2}{2}\right) = \exp\left(-\frac{k\varepsilon^2}{4}\right)$$

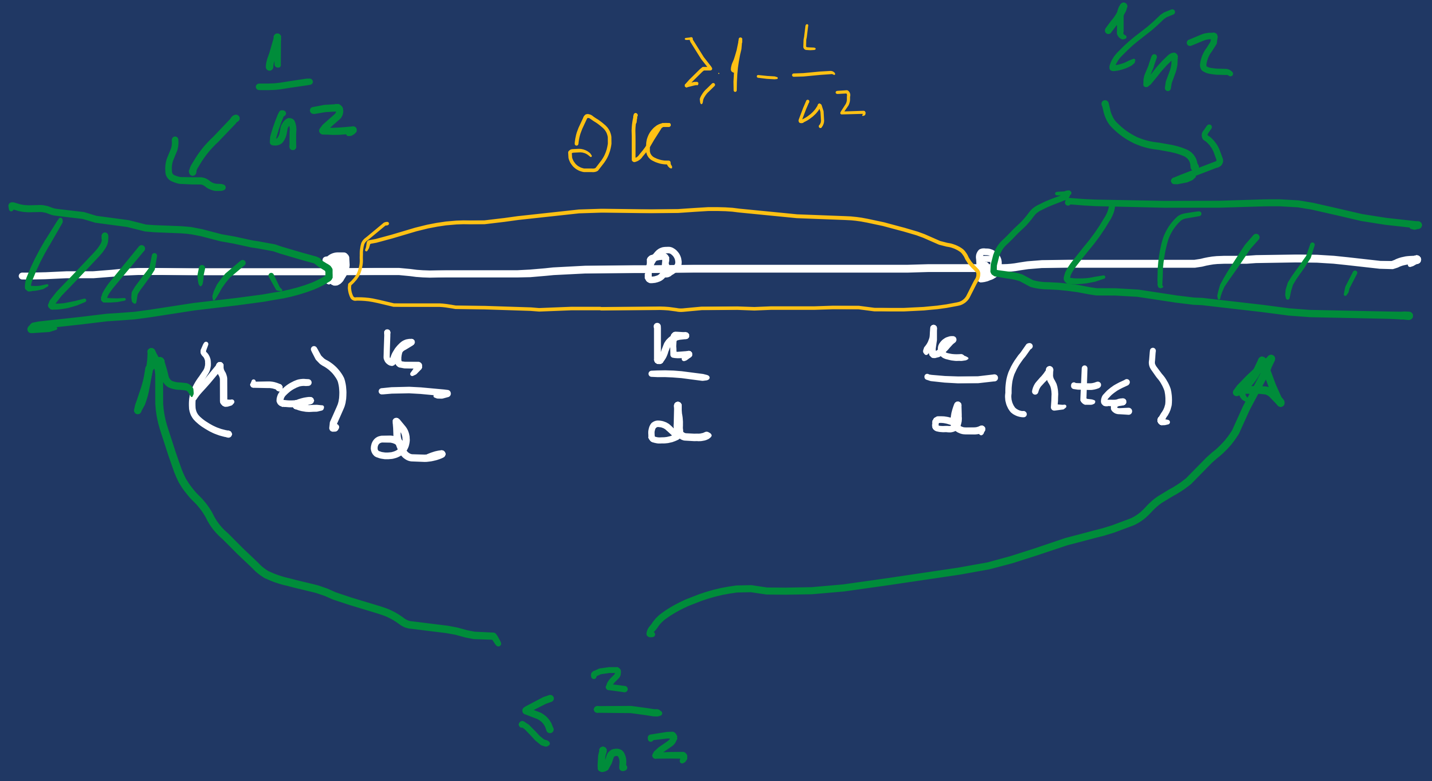
$$k \ge \frac{4}{\varepsilon^2/2-\varepsilon^3/3}\ln n \ge \frac{4}{\varepsilon^2/2}\ln n = \frac{8}{\varepsilon^2}\ln n \qquad \le \exp\left(-\frac{8}{\varepsilon^2}\ln n\,\frac{\varepsilon^2}{2}\right)$$

$$= \exp(-2\ln n) = \exp(-\ln n^2) = \frac{1}{n^2}$$

$$P\left[L \le (1-\varepsilon)\frac{k}{d}\right] \le \frac{1}{n^2} \qquad |X| = n$$

In a similar way:

$$P\left[L \ge (1+\varepsilon)\frac{k}{d}\right] \le \frac{1}{n^2} \qquad \ln(1+\varepsilon) \le \varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3},$$



This is for one pair of points.

$$P\Big[\bigvee_{1 \le i < j \le n} \neg OK_{i,j}\Big] \le \sum_{1 \le i < j \le n} \Pr[\neg OK_{i,j}] \le$$

$$\le \sum_{1 \le i < j \le n} \frac{2}{n^2} = \binom{n}{2} \frac{2}{n^2} = \frac{n(n-1)}{2} \cdot \frac{2}{n^2} =$$

$$= 1 - \frac{1}{n}.$$

$$P\Big[\bigwedge_{1 \le i < j \le n} OK_{i,j}\Big] \ge 1 - \Big(1 - \frac{1}{n}\Big) = \frac{1}{n}$$

$$\boxed{\Pr[\,S \text{ is a good subspace}\,] \geq \frac{1}{n}}$$!!

algorithm solution:

try generate a rand. plane
many times.

after $\leq n$ trials you will find
good plane.

prob. complexity: $O(n^3)$. $\qquad n\binom{n}{2} \approx \frac{1}{2}n^3$.

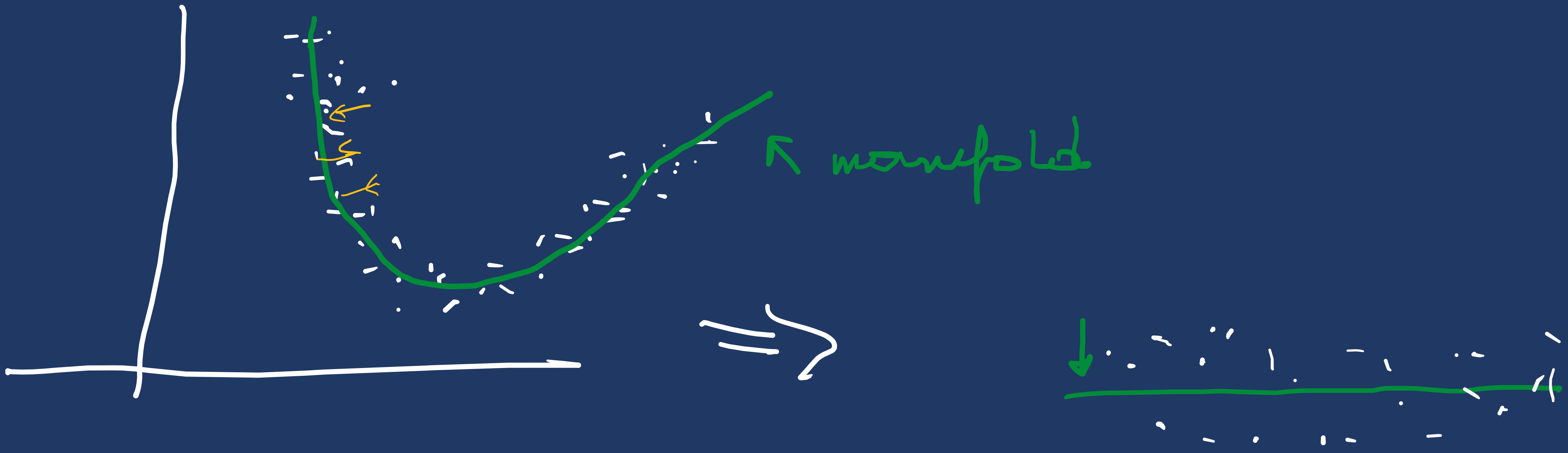
Can we do it better ?

- $k \geq \frac{4}{\varepsilon^2/2 - \varepsilon^3/3} \ln n$: essentially on require $O(\frac{1}{\varepsilon^2} \ln n)$ [$\approx$ 2005 A. Noga]
- number of trials can be reduced below $O(n^3)$ ¿ maybe $O(n^{2.1})$?

PRACTICAL SOLUTION DIM RED.

- SVD - decomp
- Princip. Comp. Analysis.

Proj. onto manifolds



nonlinear transf.