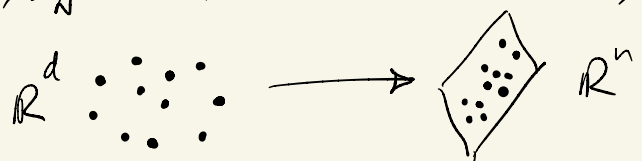
LECTURE 12

DIMENSION REDUCTION

- Data: $x_1, \dots, x_N \in \mathbb{R}^d \xrightarrow{\ ?\ } \mathbb{R}^n$, $n \ll d$?



Data compression to save on storage, speed.

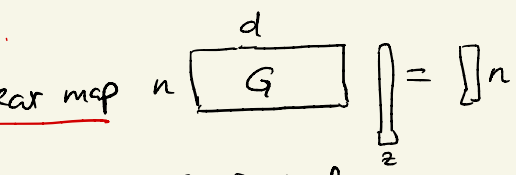
- Possible with $\boxed{n = O(\log N)}$; geometry of data preserved
  ↑ pairwise distances.

THM (Johnson–Lindenstrauss lemma '1984) $\forall\ x_1, \dots, x_N \in \mathbb{R}^d$

$\exists$ linear map $T: \mathbb{R}^d \to \mathbb{R}^n$ such that $n = C \log N$ and

$$0.99 \|x_i - x_j\|_2 \le \|T(x_i) - T(x_j)\|_2 \le 1.01 \|x_i - x_j\|_2 \quad \forall\ i,j = 1, \dots, N.$$

Proof. Probabilistic Method: → T = random projection.
choose a random linear map



$G := n \times d$ Gaussian random matrix $G_{ij} \sim N(0,1)$ iid.

FACT: $\forall$ fixed $z \in \mathbb{R}^d$, $\|z\|_2 = 1$: $Gz \sim N(0, I_n)$

$$(Gz)_i = \sum_{j=1}^{d} G_{ij} z_j \sim N\left(0, \sum_{j=1}^{d} z_j^2\right) = N(0,1) \quad \text{indep.}$$

$G_{ij} z_j \sim N(0, z_j^2)$ indep.

Thin Shell Thm ⇒

$$\mathbb{P}\{0.99\sqrt{n} \le \|Gz\|_2 \le 1.01\sqrt{n}\} \ge 1 - 2e^{-cn} \quad (*)$$

① Fix (i,j), use $(*)$ for $z = \frac{x_i - x_j}{\|x_i - x_j\|_2}$ $\Rightarrow$

$$P\left\{0.99\sqrt{n} \le \frac{\|G(x_i - x_j)\|_2}{\|x_i - x_j\|_2} \le 1.01\sqrt{n}\right\} \ge 1 - 2e^{-cn}.$$

$\Rightarrow$ for $T := \frac{1}{\sqrt{n}} G$,

$$P\{0.99\|x_i - x_j\|_2 \le \|Tx_i - Tx_j\|_2 \le 1.01\|x_i - x_j\|_2\} \ge 1 - 2e^{-cn}$$

$=: E_{ij}$

② Union Bound: $\exists\ N^2$ pairs (i,j) $\Rightarrow$

$$P\{\forall i,j : E_{ij} \text{ holds}\} \ge 1 - N^2 \cdot 2e^{-cn} = 1 - 2\exp(2\log N - cn) \quad (\ge)$$

Choose n s.t. $\boxed{cn = 4\log N}$ $\Rightarrow$

$$(\ge)\ 1 - 2\exp(-2\log N) = 1 - \frac{2}{N^2} > 0 \quad \text{if } N \ge 2.$$

(and if $N = 1$, the thm is trivial).

$\Rightarrow$ such T exists. QED.

REMARKS 1. Why does not JL lift the "curse of high dimensionality"?
Preprocess HD data by JL $\Rightarrow$ low D data.

2. Fast JL transforms. [Ailon-Chazelle '2009]

3. JL without dependence on N = #data:
Prop 9.32 of book [Liaw-Mehrabian-Plan-V' 2017]

HW: JL for Ber [Achlioptas]

COMBINATORIAL OPTIMIZATION

- **Problem A**: Given $(a_i)_{i=1}^n$, find $\max_{x_i \in \{\pm 1\}} \sum_{i=1}^n a_i x_i$

  Although exhaustive search is hard (2^n configurations of (x_i)), $\exists$ direct solution: $x_i = \text{sign}(a_i)$

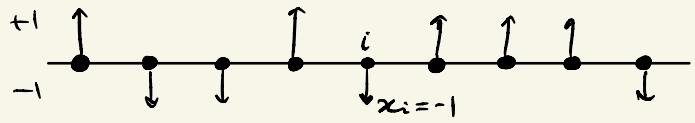
- **Problem B**: Given $(a_{ij})_{i,j=1}^n$, find

$$\max_{x_i \in \{\pm 1\}} \sum_{i,j=1}^n a_{ij} x_i x_j \qquad (*)$$

  NP-hard. Integer quadratic program.

- **EXAMPLES**:

① Ising model of magnetism:



x_i = spin of atom i (± 1)

a_{ij} = strength of interaction between atoms i,j

e.g. $a_{ij} = \begin{cases} 1 & \text{if } i,j \text{ are neighbors} \\ 0 & \text{if not} \end{cases}$

$H(x) = \sum_{ij} a_{ij} x_i x_j$ = "Hamiltonian" of the system (free energy)

- **Probability** that the system is in state $x = (x_1, \dots, x_n) \in \{\pm 1\}^n$ is

$$P(x) := \frac{1}{Z} \exp\left(-\frac{H(x)}{T}\right)$$

"Gibbs measure"

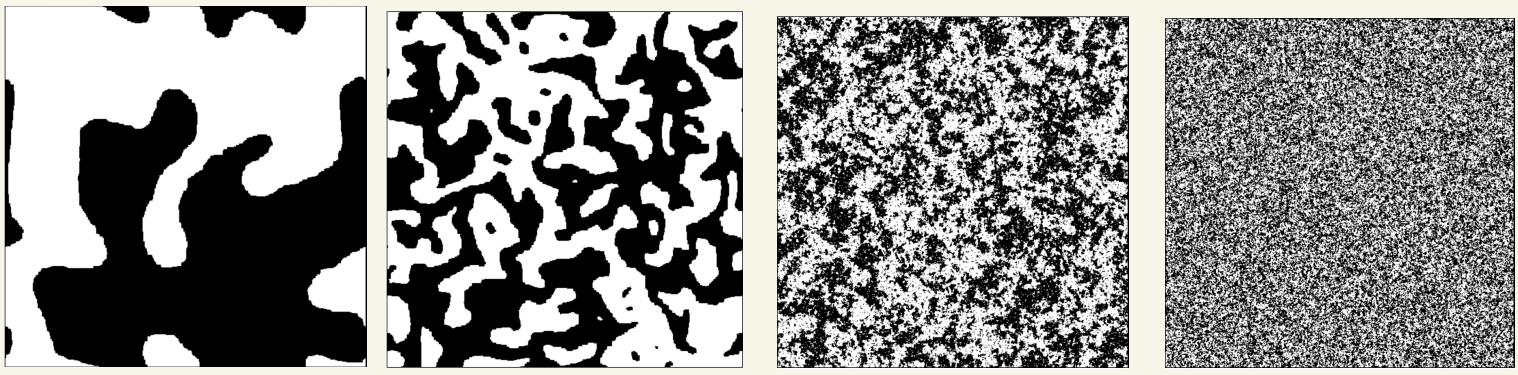
Here T = "temperature", $Z = \sum_{x \in \{\pm 1\}^n} \exp\left(-H(x)/T\right)$ is a normalizing const ("partition function")

- $T \to \infty \Rightarrow P(x) \to$ uniform on $\{\pm 1\}^n$ (chaos)

  $T \to 0 \Rightarrow P(x) \to \begin{cases} 1 & \text{for } x = x_0 \\ 0 & \text{elsewhere} \end{cases}$ where $H(x_0) = \min_{x \in \{\pm 1\}^n} H(x) = ?$

  lowest-energy state

  equivalent to Problem B.



Increasing the temperature T →