Advanced Algorithms (4232)

Lecture 1. 1/18/2022.

1) Hashing: $h: U \to [n] = \{1, 2, \dots n\}$

Dictionary problem: index set $S \subset U$

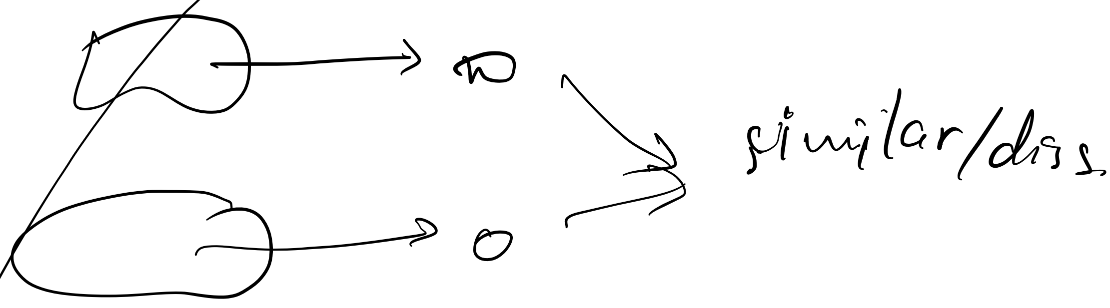
s.t. quickly $x \overset{?}{\in} S$

$O(1)$ expected time.

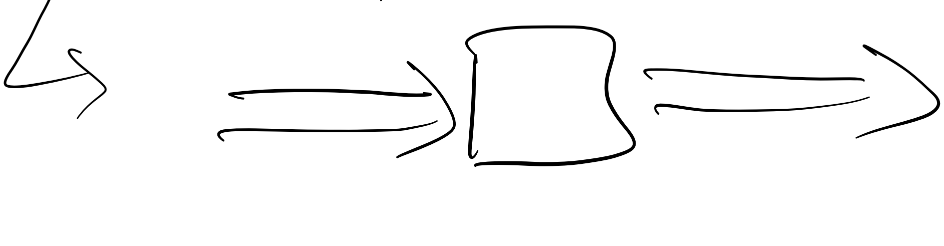
perfect hashing: $O(1)$ wor~~s~~t-case q.t.

2) Sketching / Streaming algo's.

↳ functional compression.



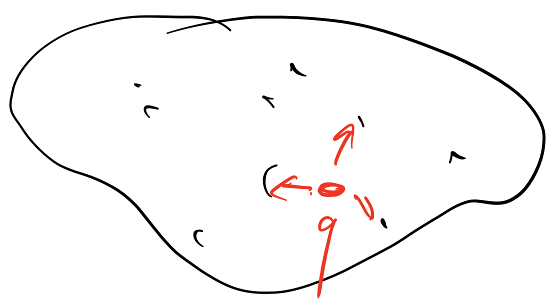
speeding up N.L.A.



3) Nearest Neighbor Search (NNS).

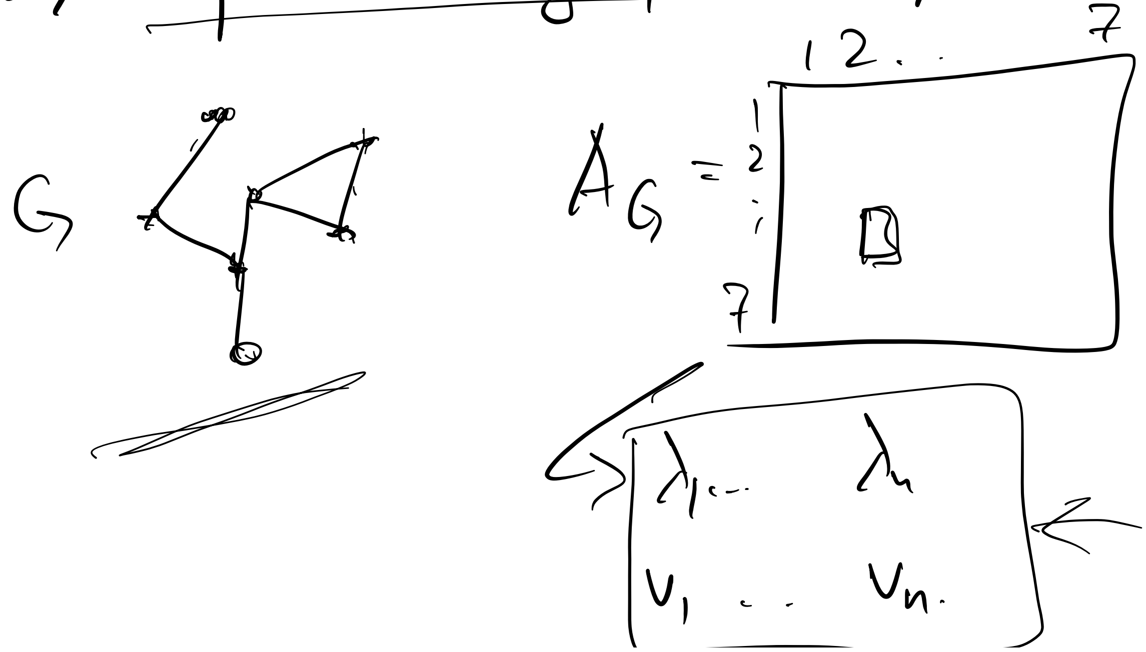
<u>problem:</u> given n points in high-d space

query: given q, report closest point from ds.



4) <u>Graph algorithms:</u> max-flow.

5) <u>Spectral graph theory.</u>



Spectral clustering

6) Optimization:

$$\min_{\substack{x \in \mathbb{R}^n \\ x \in C}} f(x)$$

LP: $f, C \rightarrow$ linear

IPM: interior-point method.

MWU: learning from experts.

7) Large-scale models.

- parallel / cluster computing.
  Map Reduce.
- I/O model: cache.

---

1) Randomized:

$\forall$ input x: $\Pr_{Alg}[\text{Alg correct}] \geq 90\%$.

2) Approximate:

$$a \le \hat{a} \le c \cdot a$$

- a ↑ correct answer
- $\hat{a}$ ↑ Alg.
- c ↑ approx ≥ 1.

1+2:

$$\Pr_{Alg}\left[a \le \hat{a} \le c \cdot a \right] \geq 90\%$$

Admin: **Pre-reqs** mathematical mat.

$O(\cdot), \Theta(\cdot)$.

Self-evaluation test **2022**

Grading:

1) HW : 5 → 50%
2) Midterm: → 25%
3) Final project. team-based
   2-4.

→ reading :

→ implementations

→ research focus.

Proposal: 1–4 pages.

Final: ~10 pages.

Policy for HWs:

- collaboration
- late:
  - 5 free late days.
  - 10% penalty /day.
  - max is 7 days. (exc. HW3)

Problem: count up to n.

Space: $\lceil \lg_2 n \rceil$ bits to store number up to n.

Thm: unless allow random and approx, can't do better.

Morris's algo: randomized & approx.

use space $O(\lg \lg n)$

$c = O(1)$. $(1+\varepsilon,\ \varepsilon = 0.1)$

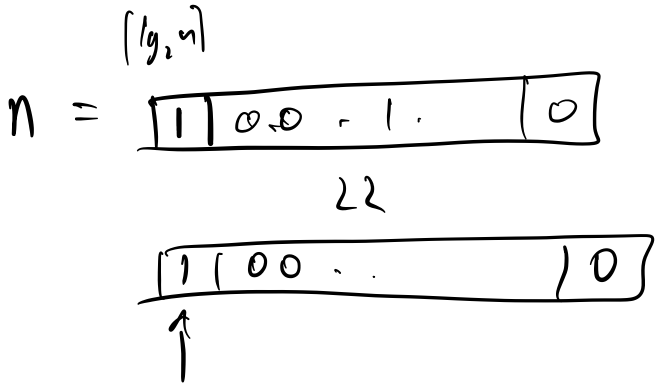
$n = 10^9$ $\lg_2 n \approx 30$

$\lg\lg n$ = single digit #s.

Algo:

- keep register X.
- init: $X = 0$.
- @ button press: $X = \begin{cases} X+1, & \text{w/ Pr} = 2^{-X} \\ X, & \text{oth.} \end{cases}$
- Estimator: $\hat{n} = 2^X - 1$.

Intuition:



X stores the index of msb of n.

Def: expectation $\mathbb{E}[X]$, X = r.v.

Fact: linearity of $\mathbb{E}$: $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$.

Def: variance: $Var[X] \triangleq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

Concentration bounds:

Markov's bound: $\forall$ r.v. $X \geq 0$:

$$\Pr[X \geq \lambda] \leq \frac{\mathbb{E}[X]}{\lambda}.$$

Chebyshev's bound: $\forall$ r.v. X:

$$\Pr\left[(X - \mathbb{E}[X])^2 \geq \lambda^2\right] \leq \frac{Var[X]}{\lambda^2}.$$