Good query performance with high probability

Up to now:

Fixed input set S_n $\quad T(S_n)$ Trapezoidal Map

$n!$ possible Query structures $D_\Pi(S_n)$ for any permutation

Any $|D_\Pi(S_n)|$ between $O(n)$ and $O(n^2)$

Query cost between $O(\log n)$ and $O(n)$

+ average size is $O(n)$
+ for fixed query point q average query cost $O(\log n)$

- Average height of $D_\Pi(S_n)$ (in any structure a long path?) worst case!
- How many $D_\Pi(S_n)$ have reasonable height (for any query point)
- Worst case query time! Construction of reasonable $D(S_n)$!

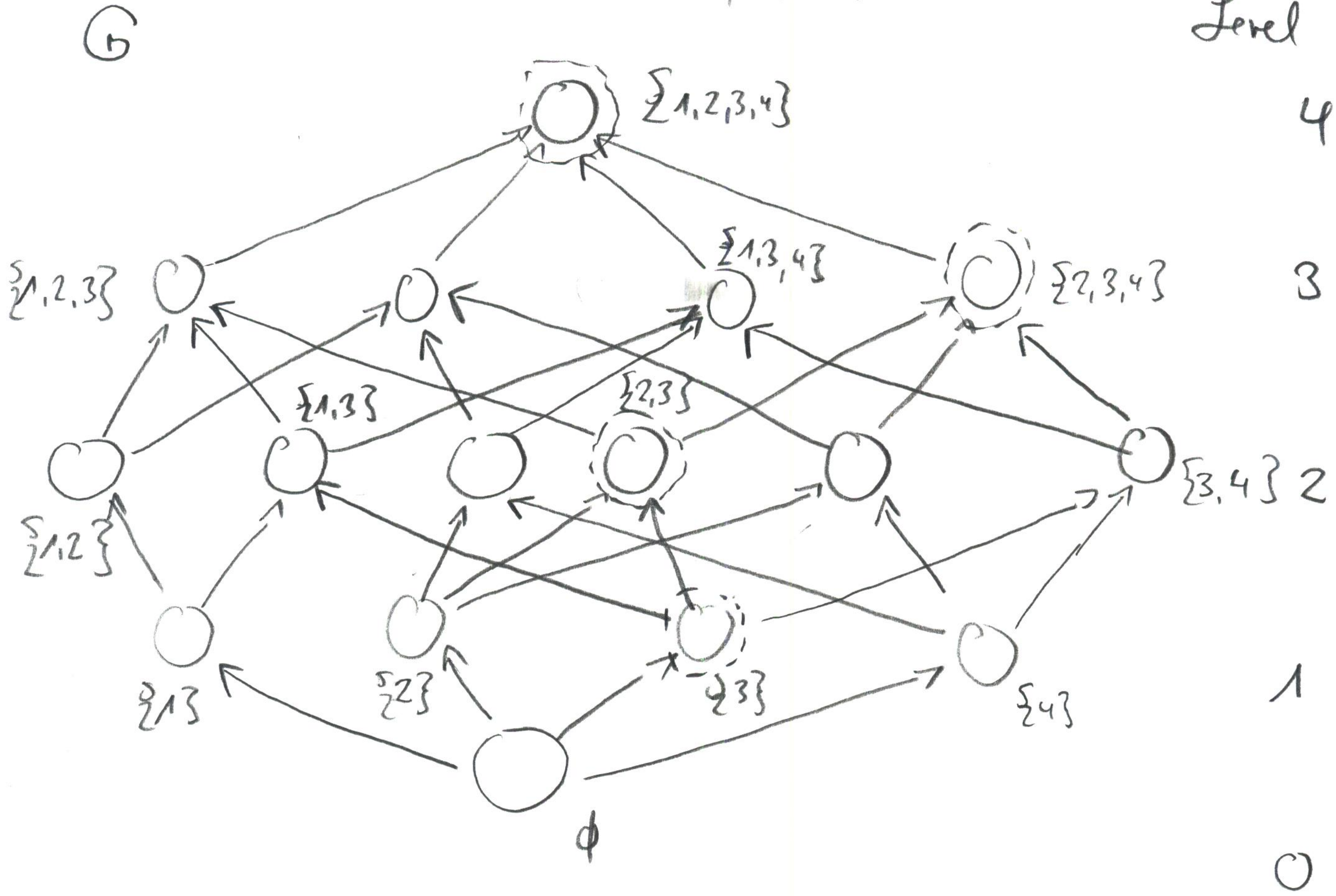
Use: Markov Inequality

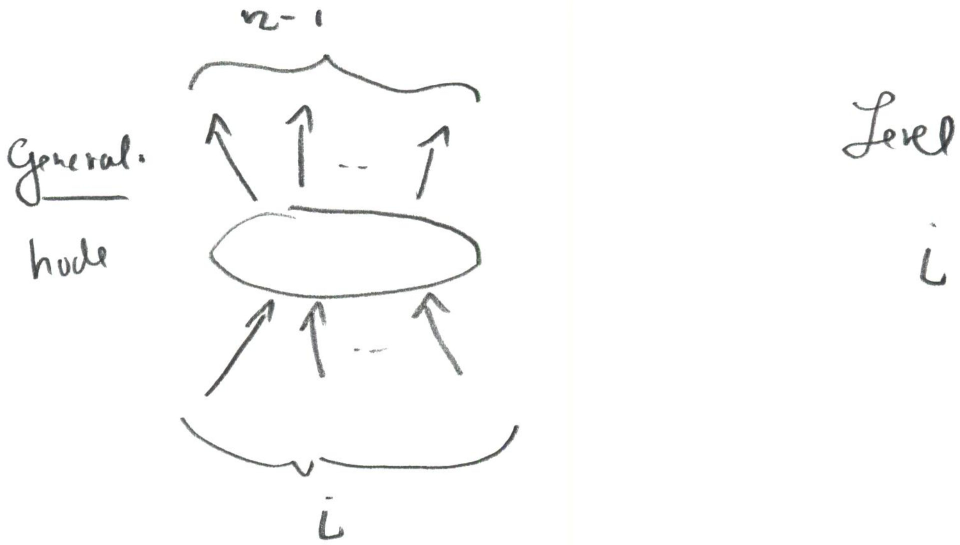
$Z \geq 0$ random variable $\alpha > 0$

$$P(Z \geq \alpha) \leq \frac{E(Z)}{\alpha}$$

Modelling with independent random variables

Construct graph that represent all insertion orders for $S_1, S_2, \ldots, S_n$
$1, 2, \ldots, n$





For any fixed insertion order there is exactly one path!

example: 3,2,4,1

n! many paths

Fixed query point q

mark edge if insertion of s changes
the current trapezoid that contains q

Max 4 line segments define the current trapezoid T

Backward analysis: At most 4 segments
change the trapezoid if they are removed from
a subset.

$\Rightarrow$ Any vertex has ≤ 4 marked edges from below

Now: If less than 4, simply mark some others arbitrarily.
(First three layers, mark all!) [Forget the reason of the marks!]

Analyze expected number of steps during which the
trapezoid containing ~~q~~ changes.
Expected number of marked edges on a source to sink path
in G.

Random variable (choose one of the $n!$ paths)

$$X_i := \begin{cases} 1 & i\text{-th edge is marked} \\ 0 & \text{otherwise} \end{cases} \quad (\text{level } i)$$

1_0 independent: mark at level i does not influence
mark at level j

"Trick" (Random path top/down!
probability to hit one of the 4 segments)

(133)

2. $P[X_i = 1] = \begin{cases} \frac{4}{i} & i \geq 4 \\ 1 & i \leq 3 \end{cases}$ probability to hit one of the four segments

3. $Y := \sum_{i=1}^{n} X_i$ # nodes on the second path to q (expected)

$3 \cdot Y$ (changes on the path to q × length 3 (add.))

Same result as before but now!

<u>Probability</u> that path has more than $O \log n$ length

$$P[Y \geq \lambda \ln(n+1)] = P[e^{tY} \geq (n+1)^{\lambda t}]$$

$\forall t > 0, \lambda > 0$

$$\geq P[\text{length search path to } q \geq 3\lambda \ln(n+1)]$$

$$\underset{\text{Markov In.}}{\leq} \frac{E(e^{tY})}{(n+1)^{\lambda t}}$$

$$\underset{E(e^{tY}) = n+1 \text{ for } t := \ln \frac{5}{4}}{\leq} \frac{n+1}{(n+1)^{\lambda t}} = \frac{1}{(n+1)^{\lambda t - 1}}$$

$$\mathbb{E}\left(e^{tY}\right) = \mathbb{E}\left(e^{t\sum_{i=1}^{n} X_i}\right)$$

$$= \mathbb{E}\left(\prod_{i=1}^{n} e^{tX_i}\right)$$

$$= \prod_{i=1}^{n} \mathbb{E}\left(e^{tX_i}\right)$$

$\uparrow$ indep. var.

[e^t with prob. $\frac{4}{i}$, e^0 with prob. $\left(1-\frac{4}{i}\right)$]

$$= \prod_{i=1}^{n} e^{t}\cdot\frac{4}{i} + e^{0}\left(1-\frac{4}{i}\right)$$

$\uparrow$

$$\left(\mathbb{E}(X_i) = 1\,\frac{4}{i} + 0\cdot\left(1-\frac{4}{i}\right)\right)$$

$$= \prod_{i=1}^{n} \frac{5}{4}\,\frac{4}{i} + \frac{i-4}{i}$$

$$= \prod_{i=1}^{n} \frac{i+1}{i} = \frac{2}{1}\,\frac{3}{2}\cdots\frac{n+1}{n}$$

$$= n+1$$

Lemma 50 Let S be a set of n non-crossing line segments, let q be a query point and λ be a parameter $\lambda > 0$. The probability that the search path for q in $\mathcal{D}(S_n)$ (rand. increm. constr.) has more than $3\lambda \ln(n+1)$ steps is at most $\frac{1}{(n+1)^{\lambda \ln \frac{5}{4} - 1}}$.

Proof: Just shown!

Statement for single search path!

Maximum length of a search path?

Height of $\mathcal{D}_{\overline{N}}(S_n)$?

Problem too many query points!

Collect points with the same search path in $\mathcal{D}_{\overline{N}}$

Slab / Strip method: vertical lines through vertices intersected by line segments

at most $2(n+1)^2$ cells

For any cell the search path is the same for all points in the cell!

$$P\left[\text{height}\left(D_{\bar{\pi}}(S_n)\right) \geq 32\ln(n+1)\right]$$

$$\leq P\left[\text{one of the } 2(n+1)^2 \text{ search paths has height} \geq 32\ln(n+1)\right]$$

$$\leq 2(n+1)^2\, P\left[\text{fixed search path has height} \geq 32\ln(n+1)\right]$$

$$\underset{\substack{\uparrow \\ \text{Lemma 50}}}{\leq} \frac{2(n+1)^2}{(n+1)^{2\ln\frac{5}{4}-1}} = 2\,\frac{1}{(n+1)^{2\ln\frac{5}{4}-3}}$$

<u>Lemma 51</u> Same precondition as in Lemma 50.

The probability that the maximum length of a search path in $D(S_n)$ (rand. ins. constr.) has length greater than $32\ln(n+1)$ is at most

$$\frac{2}{(n+1)^{2\ln\frac{5}{4}-3}}.$$

Proof: Just given!

<u>Consequence</u>: Choose $2 := 20$ $\qquad \ln\frac{5}{4} \approx 0.223$

$$\Rightarrow P\left[\text{height}\left(D(S_n)\right) \geq 60\ln(n+1)\right] \leq \frac{2}{(n+1)^{1.4}} < \frac{1}{4}$$

for $n > 4$.

$$\Rightarrow \quad P[\text{height}(D(S_n)) \leq 60 \ln(n+1)] \geq \frac{3}{4}$$

Similar arguments: $P[\text{size}(D(S_n)) \leq 2n] \geq \frac{3}{4}$

$$\Rightarrow \quad P[\text{Construction cost}(D(S_n)) \leq 2n \log n] \geq \frac{3}{4} \cdot \frac{3}{4} \geq \frac{1}{2}$$

(max. length of search path
max size)

Similar arguments: also $P[\text{Constr. cost}(D(S_n)) \leq 2n \log n] \geq \frac{3}{4}$

Theorem 52: S planar subdivision of n edges.
There exists a point location data structure for S of size $O(n)$ with $O(\log n)$ query time. Can be computed in $O(n \log n)$ expected time.

Proof: Build: Choose random permutation π

Build: $D_\pi(S_n)$

If choice is bad, start again! $\quad (\text{height}, \text{size}) > (C \cdot \log n, C \cdot n)$

In the average 4 attempts are enough