Supported by IDEI Grant PN-II-ID-PCE-2011-3-0981 ”Structure and computational difficulty in combinatorial optimization: an interdisciplinary approach”.

Heapable subsequences and related concepts

Gabriel Istrate, Cosmin Bonchiş

West University of Timişoara
and the e-Austria Research Institute

http://tcs.ieat.ro

Introduction

Starting Point: Longest Increasing Sequence

3 2 5 7 1 6 9

- Given n (integer) numbers $a_1, a_2, \ldots, a_n$ find the longest subsequence (not necessarily contiguous) that is increasing
- First-year algorithms: Dynamic programming.
- Another (greedy, also first-year) algorithm: Patience sorting.

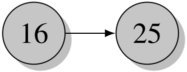
Start (greedily) building decreasing piles. When not possible, start new pile.

Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20

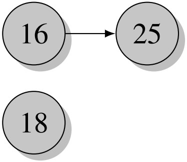
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



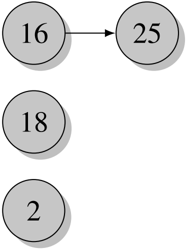
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



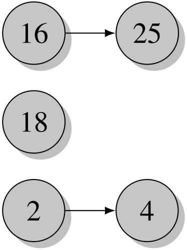
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



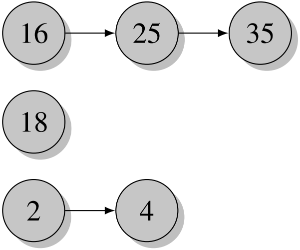
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



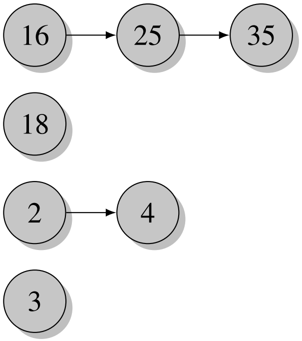
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



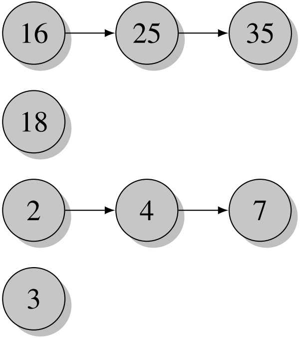
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



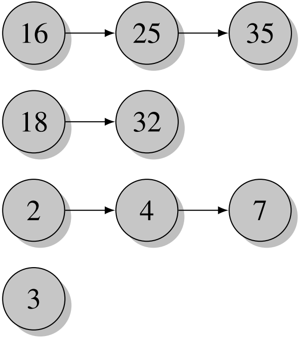
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



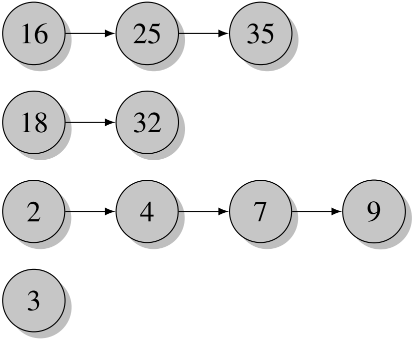
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



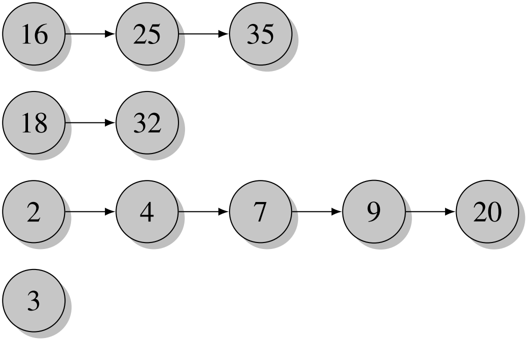
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



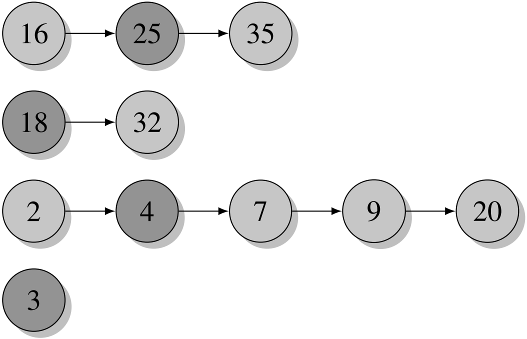
Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



Patience sorting

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20

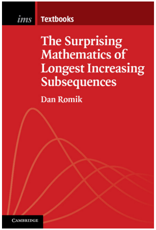


- Partitions the array into increasing (up-)sequences
- SUS(A): the minimal number of such subsequences.
- THEOREM(classical): $SUS(A) = LDS(A)$.

Longest increasing sequence of a random permutation

$$E_{\pi \in S_n}[LIS(\pi)] = 2\sqrt{n} \cdot (1 + o(1)).$$

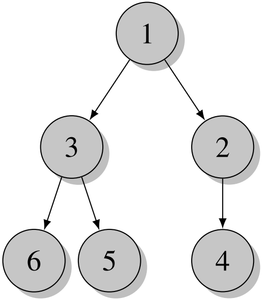
- Logan-Shepp (1977), Veršik-Kerov (1977), Aldous-Diaconis (1995)
- Very rich problem. Connections with nonequilibrium statistical physics and theory of Young tableaux



From data structures to patterns in sequences

Byers, Heeringa, Mitzenmacher, Zervas ANALCO'2011

Sequence of integers A is heapable if it can be inserted into binary heap-ordered tree (not necessarily complete) without having to call HEAPIFY.



Example: 1 3 2 6 5 4   Counterexample: 5 1 . . .

Byers et al. results on heapability

- Polynomial time algorithm for heapability.
- complete heapability: NP-complete
- Longest Heapable Subsequence (LHS): complexity open.
- But with high probability $LHS(\pi) = n - o(n)$, where $\pi \in S_n$ is a random permutation.

Intuition: heapability "weak" versions of increasing sequences.
Recall 1 3 2 6 5 4.
Question (Byers et al.): Does the theory of LIS extend to heapable sequences ?

”Patience heaping”

k-heapable: heapable into a *k*-ary min-heap

$MHS_k(A)$: Extension of SUS. the smallest number of *k*-heapable subsequences in a decomposition of A.

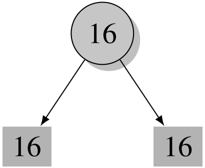
Slot of a node: the k (free) positions where children may grow.

Algorithm 1.1: PATIENCE-HEAPING(W)

INPUT $W = (w_1, w_2, \ldots, w_n)$ a list of integers.
Start with empty heap forest $T = \emptyset$.
for i in range(n):
  if (there exists a slot where X_i can be inserted):
    insert X_i in the slot with the lowest value.
  else :
    start a new heap consisting of X_i only.

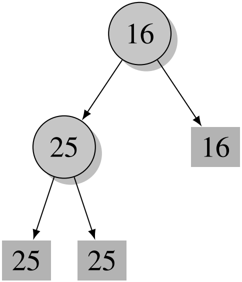
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



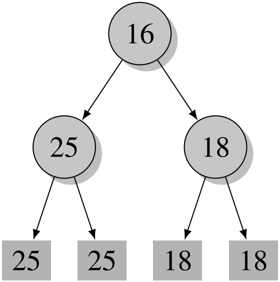
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



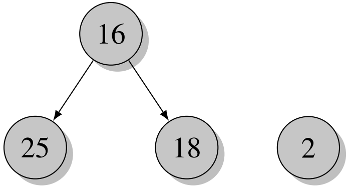
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



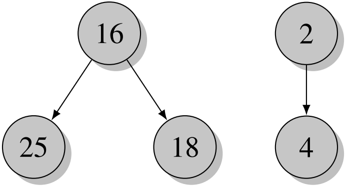
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



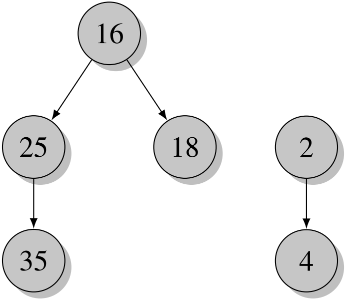
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



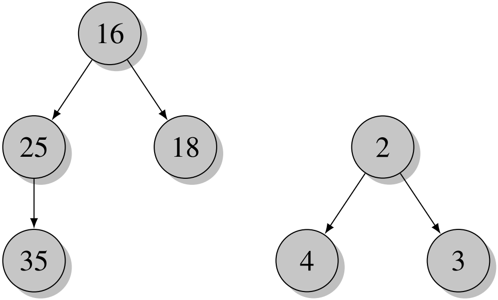
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



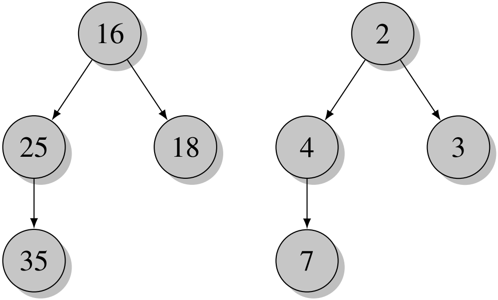
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



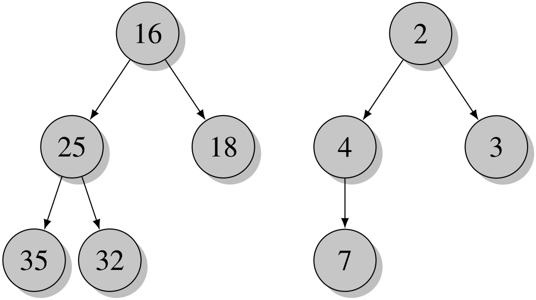
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



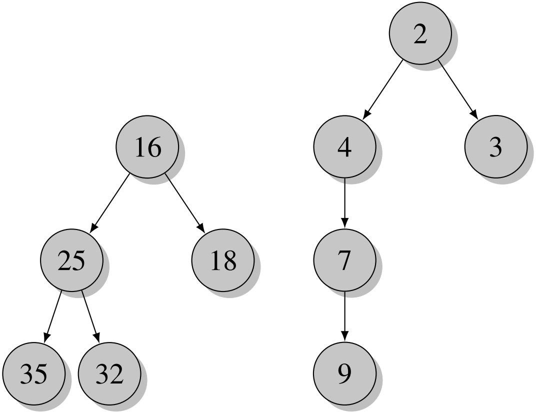
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



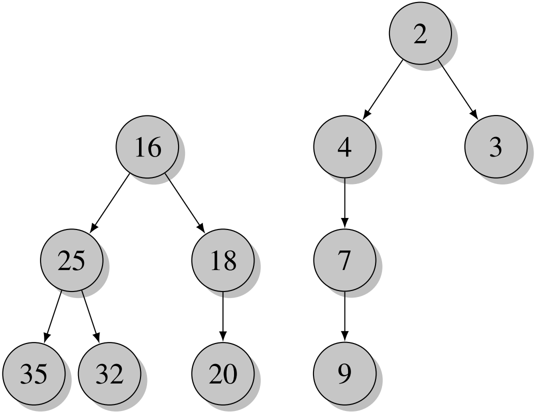
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



Two (easy) results

THEOREM 1: ”Patience heaping” computes $MHS_k(A)$.
THEOREM 2: (a). there exists sequences X such that
(a). $MHS_k(X) < MHS_{k-1}(X) < \ldots < MHS_1(X)$.
(b). $\sup_X [MHS_{k-1}(X) - MHS_k(X)] = \infty$.

Proof Ideas:

- (a). Define domination relation between multisets of slots.
- Greedy insertion dominates any other insertion+ induction.

  If GREEDY creates new heap then any other algorithm does.

- (b). Thm.1 + Example.

Scaling of MHS_k

How does $E[MHS_k(\pi)]$, where π is a random in S_n, behave ?

- Increasing $\equiv$ "1-heapable". $E[LIS(A)] \sim 2\sqrt{n}$.
- For any k growth at least logarithmic.

THEOREM 3: For every fixed $k, n \geq 1$

$$E_{\pi \in S_n}[MHS_k(\pi)] \geq H_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}, \quad (1)$$

the n'th harmonic number, $\sim \ln(n)$.

Proof Idea:
Sequence minima start new heaps.

A beautiful conjecture

CONJECTURE: We have

$$\lim_{n\to\infty} \frac{E[MHS_2[\pi]]}{\ln(n)} = \phi,$$

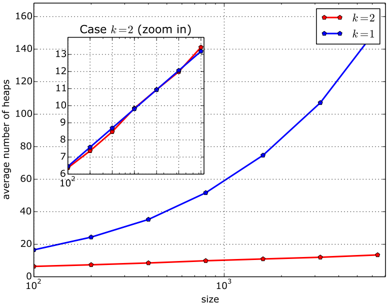
with $\phi = \frac{1+\sqrt{5}}{2}$ the golden ratio.

More generally, for an arbitrary $k \geq 2$ the relevant scaling is

$$\lim_{n\to\infty} \frac{E[MHS_2[\pi]]}{\ln(n)} = \frac{1}{\phi_k}, \tag{2}$$

where ϕ_k is the unique root in $(0, 1)$ of equation $X^k + X^{k-1} + \ldots + X = 1$.

Evidence: Simulations.



- We kind of know what's going on.
- Can make nonrigorous computations that match experimental predictions
- . . . like physicists do !

We just can't make all steps of the argument rigorous !

Hammersley's process: an interacting particle system.

Tops of piles in patience sorting = live particles in Hammersley's process:

- Particles arive at integer times as random real numbers $X_i \in [0, 1]$.
- Particle X_i kills closest live particle X_j, $X_i < X_j$ (if any)

- studied in the area of interacting particle systems, a field between probability theory and (Nonequilibrium) Statistical Physics.
- relative of a more famous process, the so-called Totally Asymmetric Exclusion Process (TASEP)
- Most illuminating proof of $E[LIS(\pi)] \sim 2\sqrt{n}$ (Aldous-Diaconis) analysis of the so-called hydrodynamic limit of Hammersley's process.

Physics of patience heaping ?

Hammersley's process with k lifelines (HAM$_k$):

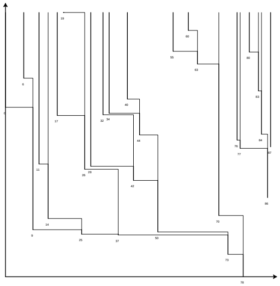
- "Particles" = random numbers $X_i \in [0, 1]$.
- each "particle" is initially endowed with k lives.

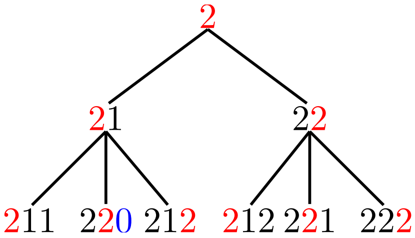
- X_i ~~kills~~ takes one lifeline from closest live X_j, $X_i < X_j$ (if any)

THEOREM 4: At time n the multiset of free slots in patience heaping corresponds (with multiplicities) to live particles in Hammersley's process with k lifelines.

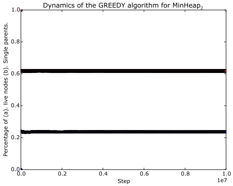
Combinatorial view of process HAM_k

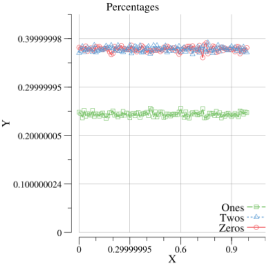
- Words over alphabet $0, 1, 2$ and a (conventional leading) -1.
- Start with $W_0 = -1$ (leftmost marker; not shown below)
- Choose a random position to the right of -1. Put there a 2. Remove 1 from the closest nonzero digit to the right (if any).
- # of heaps = # of insertions that don't remove a lifeline.





A ”physicist” explanation for the dynamics of HAM_k





Exp: 5 indep. runs, 10.000.000 steps. Final vals: Fig 1. First trajectory: Fig 2, once every 10.000 steps. Fig 3: binned densities.

- 1. $n \to \infty$: Limit of W_n = compound Poisson process. W_n = random string of 0,1,2 (densities d_0, d_1, d_2).
- 2. Assuming well mixing of digits one can write evolution equations that predict (constant) limit values of d_0, d_1, d_2.
- 3. From this: ”on average” the probability that the number of heaps increases at stage $n \sim \frac{1+\sqrt{5}}{2 \cdot n}$.

Rigorous result on HAM$_k$

$d_0(n), d_1(n), d_2(n)$: average densities of digits 0,1,2 in word W_n (discarding -1).

FIRST STEP: There exist constants $d_0, d_1, d_2 \in [0, 1]$ such that

$$\lim_{n\to\infty} d_i(n) = d_i, i = 0, 1, 2.$$

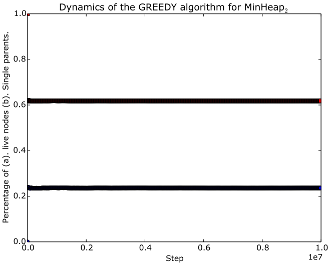
Tool: subadditivity/Fekete's Lemma: "If a_n is a sequence with $a_{m+n} \leq a_m + a_n$ then $\lim_{n\to\infty} a_n/n$ exists."

In the paper: Fekete's lemma applies to some suitable independent linear combinations of $d_i(n)$.

A "physicist" explanation for the dynamics of HAM_k

Empirically $d_1 + d_2 = \frac{\sqrt{5}-1}{2} \sim 0.618\ldots$, $d_2 = \sqrt{5} - 2 \sim 0.236$ and of course $d_0 + d_1 + d_2 = 1$.

Five independent runs, each with 10.000.000 simulation steps. Final values: Figure 1. First trajectory: Figure 2, only once every 10.000 steps (we make 10.000.000 moves !). We've converged very fast.



A physicist like's explanation for the dynamics of HAM_k (III)"

- Assuming well mixing, largest nonzero digit (there are $\sim \frac{\sqrt{5}-1}{2}n$ of them) "has rank $\sim (n - \frac{\sqrt{5}+1}{2})$"
- The probability that the new particle at time n come above this element, therefore increasing the number of heaps, is $\sim \frac{\sqrt{5}+1}{2n}$.
- Thus $E[MHS_2(\pi)]$ is $\sim \frac{\sqrt{5}+1}{2} \cdot H_n \sim \frac{\sqrt{5}+1}{2} \ln(n)$.

Will explain/substantiate all these in a subsequent, "physics-like" paper.