Cyclic Increasing Sequences

Michael Albert [1] Mike Atkinson[1] Doron Nussbaum[2] Jörg Sack[2] Nicola Santoro[2]

[1]Department of Computer Science, University of Otago

[2]School of Computer Science, Carleton University

ACCMCC'30, December 2005

Outline of talk

1. Longest increasing subsequences
2. Finding the LIS
3. The circular problem
4. Analysis

Increasing subsequences

- A sequence of numbers: 4 3 1 6 2 5 7

Increasing subsequences

- A sequence of numbers: 4 3 1 6 2 5 7
- An increasing subsequence: 4 3 1 6 2 5 7

Increasing subsequences

- A sequence of numbers: 4 3 1 6 2 5 7
- An increasing subsequence: 4 3 1 6 2 5 7
- A longer increasing subsequence: 4 3 1 6 2 5 7

Increasing subsequences

- A sequence of numbers: 4 3 1 6 2 5 7
- An increasing subsequence: 4 **3** 1 **6** 2 5 7
- A longer increasing subsequence: 4 **3** 1 **6** 2 5 **7**
- The longest increasing subsequence: 4 3 **1** 6 **2** **5** **7**

How Long?

- $1\ 2\ 3\ \cdots\ n$: LIS has length n

How Long?

- $1\ 2\ 3\ \cdots\ n$: LIS has length n
- $n\ n-1\ \cdots 2\ 1$: LIS has length 1

How Long?

- $1\ 2\ 3\ \cdots\ n$: LIS has length n
- $n\ n-1\ \cdots 2\ 1$: LIS has length 1
- What is typical?

How Long?

- $1\ 2\ 3\ \cdots\ n$: LIS has length n
- $n\ n-1\ \cdots 2\ 1$: LIS has length 1
- What is typical?
- (Ulam, 1960's) Given a random arrangement of $1, 2, \ldots, n$ what is the expected length of its longest increasing subsequence?

The expected length of the LIS

Notation:

n length of a random sequence, μ the expected length of its LIS.

- Erdös and Szekeres (1935): $\mu \geq \frac{1}{2}\sqrt{n}$

The expected length of the LIS

Notation:

n length of a random sequence, μ the expected length of its LIS.

- Erdös and Szekeres (1935): $\mu \geq \frac{1}{2}\sqrt{n}$
- Baer and Brock (1968): guessed $\mu \sim 2\sqrt{n}$

The expected length of the LIS

Notation:

n length of a random sequence, μ the expected length of its LIS.

- Erdös and Szekeres (1935): $\mu \geq \frac{1}{2}\sqrt{n}$
- Baer and Brock (1968): guessed $\mu \sim 2\sqrt{n}$
- Logan and Shepp (1977): proved $\mu \sim 2\sqrt{n}$

The expected length of the LIS

Notation:

n length of a random sequence, μ the expected length of its LIS.

- Erdös and Szekeres (1935): $\mu \geq \frac{1}{2}\sqrt{n}$
- Baer and Brock (1968): guessed $\mu \sim 2\sqrt{n}$
- Logan and Shepp (1977): proved $\mu \sim 2\sqrt{n}$
- Baik, Deift, and Johansson (1999): found the complete distribution – standard deviation σ and all.

The expected length of the LIS

Notation:

n length of a random sequence, μ the expected length of its LIS.

- Erdös and Szekeres (1935): $\mu \geq \frac{1}{2}\sqrt{n}$
- Baer and Brock (1968): guessed $\mu \sim 2\sqrt{n}$
- Logan and Shepp (1977): proved $\mu \sim 2\sqrt{n}$
- Baik, Deift, and Johansson (1999): found the complete distribution – standard deviation σ and all.
- $$\begin{aligned} \mu &= 2n^{1/2} - \gamma n^{1/6} + o(n^{1/6}) \\ \sigma &= \delta n^{1/6} + o(n^{1/6}) \end{aligned}$$

  where $\gamma = 1.711$ and $\delta = 0.902$.

The LIS algorithm: Schensted (1961)

- Scan from left to right
- Maintain “best” IS’s of each length
- Update the best IS’s after reading each term
- Ultimately return the longest of the best IS’s

What does “best” mean?
Most easily extended – i.e. having smallest last term.
Often just need the final term of a best IS.

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25											
2												
3												
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25										
2		30										
3												
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18									
2		30	30									
3												
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18								
2		30	30	20								
3												
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18							
2		30	30	20	20							
3					56							
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18						
2		30	30	20	20	20						
3					56	21						
4												
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18					
2		30	30	20	20	20	20					
3					56	21	21					
4							45					
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17				
2		30	30	20	20	20	20	20				
3					56	21	21	21				
4							45	45				
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17	17			
2		30	30	20	20	20	20	20	20			
3					56	21	21	21	21			
4							45	45	40			
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17	17	15		
2		30	30	20	20	20	20	20	20	20		
3					56	21	21	21	21	21		
4							45	45	40	40		
5												

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17	17	15	15	
2		30	30	20	20	20	20	20	20	20	20	
3					56	21	21	21	21	21	21	
4							45	45	40	40	40	
5											43	

Example: LIS of 25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17	17	15	15	15
2		30	30	20	20	20	20	20	20	20	20	16
3					56	21	21	21	21	21	21	21
4							45	45	40	40	40	40
5											43	43

Example: LIS of $25, 30, 18, 20, 56, 21, 45, 17, 40, 15, 43, 16$

	25	30	18	20	56	21	45	17	40	15	43	16
1	25	25	18	18	18	18	18	17	17	15	15	15
2		30	30	20	20	20	20	20	20	20	20	16
3					56	21	21	21	21	21	21	21
4							45	45	40	40	40	40
5											43	43

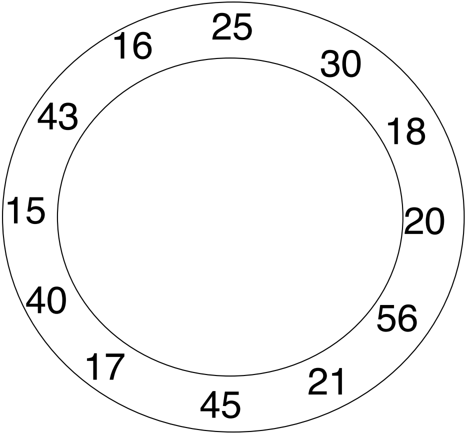
The LIS has length 5 and is $18, 20, 21, 40, 43$

Analysis

- Only store the last column
- New last column obtained from old by binary search and insertion
- Some back pointers to actually construct the result
- Total time $O(n \log n)$

The Longest Increasing Circular Subsequence (LICS)

$25, 30, 18, 20, 56, 21, 45, 18, 40, 15, 43, 16$ in a circle.



The LICS is no longer $18, 20, 21, 40, 43$ but is $15, 16, 18, 20, 21, 45$

How do we find the LICS?

- We could run the LIS algorithm on the n different ways of regarding the circular sequence as a linear sequence.
- That takes time $O(n^2 \log n)$
- Can we do better?

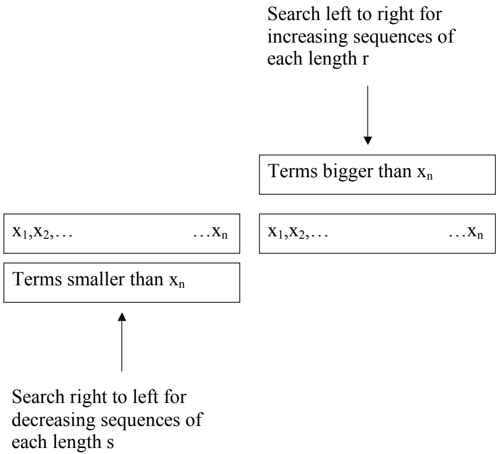
Passing through

Lemma

If $t_1, \ldots, t_r$ is the final column of the LIS algorithm then there is an LICS that contains at least one of these terms.

There and back again

We can find the LICS in time $O(n \log n)$ if we know any term that it contains.



Algorithm for the LICS

- Run the LIS algorithm and find the last column $t_1, \ldots, t_r$
- Find the longest circular sequence through each t_i
- Return the longest of these

The execution time

Theorem

The expected execution time of the LICS algorithm is $O(n\sqrt{n}\log n)$.

The execution time

Theorem

The expected execution time of the LICS algorithm is $O(n\sqrt{n}\log n)$.

- There is also a variant which has worst case execution time $O(n\sqrt{n}\log n)$ with a vanishing probability of getting the wrong answer.

The execution time

Theorem

The expected execution time of the LICS algorithm is $O(n\sqrt{n}\log n)$.

- There is also a variant which has worst case execution time $O(n\sqrt{n}\log n)$ with a vanishing probability of getting the wrong answer.
- All the implied O constants are small.

Lies, damned lies and statistics

10000 random trials gave these estimates μ-est. for μ for various n.

n	10000	12000	14000	16000	18000	20000
μ-est.	200.1	219.3	237.0	253.4	268.8	283.4

Lies, damned lies and statistics

10000 random trials gave these estimates μ-est. for μ for various n.

n	10000	12000	14000	16000	18000	20000
μ-est.	200.1	219.3	237.0	253.4	268.8	283.4
$2\sqrt{n}$	200.0	219.1	236.7	253.0	268.3	282.8

It appears that μ is much closer to $2\sqrt{n}$ than it was in the linear case. But all we can prove is. . .

The mean of the LICS

Theorem

$$\lim_{n \to \infty} \frac{\mu(n)}{2\sqrt{n}} = 1.$$