OPEN PROBLEMS in the ANALYSIS of SORTING and SEARCHING ALGORITHMS

ROBERT SEDGEWICK
Princeton University

Warmup: a combinatorial puzzle

Average number of inversions in
a random 2-ordered permutation

.	1	2			0		1	2	3	4	5	6	0		1	3	2	4	5	6	1
.	2	1			1		1	2	3	4	6	5	1		1	3	2	4	6	5	2
.							1	2	4	3	5	6	1		1	2	3	5	4	6	1
.	1	2	3	4	0		1	2	4	3	6	5	2		1	3	2	5	4	6	2
.	1	2	4	3	1		2	1	3	4	5	6	1		1	4	2	5	3	6	3
.	2	1	3	4	1		2	1	3	4	6	5	2		2	1	3	5	4	6	2
.	2	1	4	3	2		2	1	4	3	5	6	2		2	1	5	3	6	4	4
.	3	1	4	2	3		2	1	4	3	6	5	3		3	1	5	2	6	4	5
.	1	3	2	4	1		3	1	4	2	5	6	3		4	1	5	2	6	3	6
.							3	1	4	2	6	5	4		1	2	5	3	6	4	3

Cumulative counts
inversions in all 2-ordered perms of size 2N

.	N	1	2	3
.	C(N)	1	8	48
.	C(N)/N	1	4	16

C(N) =

Simple proof?

Average number of inversions
in a 2-ordered permutation

Inversions in 2-ordered Permutations

Correspondence with random walks, random trees

- 2-ordered perm:
  - random walk starting and ending at 0
- number of inversions:
  - area under the curve

Direct combinatorial sum

Elementary BGF derivation

Is there a one-line proof??

- Ex: Divide-and-conquer recurrence?
  - C(N+1) = 4C(N) + 4^N
- Ex: Direct correspondence with bitstrings?
  - 4*C(N) = # bits in all (2N)-bit numbers

Hashing with linear probing

N keys, table size M

hash function h(K)

put key in first unoccupied
position among h(K), h(K)-1, h(K)-2 ...

average cost to insert N keys?

Knuth volume 3 asterisk, 1962

intricate argument yields triple sum

evaluate sum with Abel's binomial theorem

$$C(N+1, M) = (1/2)(N/M + (N/M)((N-1)/M) + ...)$$

BGF derivation, 1997

Analysis of random maps

Knuth, to appear

OPEN PROBLEM: solution satisfies

$$C(N+1, M) = (N/M)(C(N, M) + 1/2)$$

Simple proof??

Shellsort

```
shellsort(itemType a[], int l, int r)
 {
    int incs[16] =
       { 1391376, 463792, 198768, 86961, 33936,
          13776, 4592, 1968, 861, 336, 112, 48,
          21, 7, 3, 1 };
    int i, j, h, v;

    for ( k = 0; k < 16; k++)
       for (h = incs[k], i = l+h; i <= r; i++)
          {
             v = a[i]; j = i;
             while (j > h && a[j-h] > v)
               { a[j] = a[j-h]; j -= h; }
             a[j] = v;
          }
 }
```

Running time depends on increment sequence

Worst case

* upper bound
* lower bound

Average case

* OPEN

Goal: find the best increments through analysis

www.cs.princeton.edu/~rs/shell

Average-case analysis of Shellsort

Even the simplest cases are difficult to analyze

Two increments: (h, 1) Shellsort

Three increments: (h, k, 1) Shellsort

Account for dependence on number-theoretic
properties of increments (Frobenius problem)?

Best increments??

Variants of Shellsort
- h-bubble (Dobosiewicz)
- h-shake (Incerpi-Sedgewick)
- h-brick (Sedgewick, Lemke)

O(log N) probabilistic sorting networks?

Balanced search trees

BSTs

AVL trees
top-down RB trees
B-trees
splay trees
skip lists

All encompassed in the same general scheme
* one bit per node
* perform rotation operations to balance trees

Worst case: O(log N) guaranteed search cost
Average case:
Are balanced trees asymptotically optimal?
(~lg N path length with coefficient 1)
OPEN

Approaches to balanced tree analysis

Fringe analysis

- Markov process at bottom of tree
- provides upper bound on path length
- chain size exponential in number of levels
- no chance to get bound to 1

Tree enumeration

- T(z) = T(z^2 + z^3)
- oscillating leading term
- no real connection to search tree problem

"Top-down" trees

- algebraic structure reflected in GFs?

Open problems, 1972 (Knuth volume 3)

Problems rated 46-50 considered OPEN

55 such problems

19 Sorting algorithms

18 Networks and lower bounds

18 Searching algorithms

Scorecard for 1997

SOLVED: 19

NEW: 4

TO DO: 40

SORTING (Knuth 1972 open problems)

Avg. length of longest increasing subsequence
Generalize tableaux to 3D
Shellsort worst case
Average case for Batcher’s sort
Average number of passes for shaker sort
Variance for selection sort quantity
Heapsort average case
Enumerate leftist trees
Stack sorting algorithm
Analyze replacement selection
Polyphase optimal minipass?
Minimal path length in T-way merging trees
N log N required to reverse records on tape?
Minimize phases in tape merging
Tape merging with buffers
Polynomial algorithm for optimal disk merge trees?
Uniform optimal merge trees?
Lower bound for optimal merge trees
Stable sorting in minimal storage

NETWORKS/LOWER BOUNDS (Knuth 1972 problems)

Exact value of S(N) for infinitely many N
Hwang-Lin merging
Convexity properties of M(m,n)
Minimean merging
Minimean selection
Minimax selection third largest
Minimax median
Minimean median
Networks with m-sorter modules
Beat odd-even for merging (comparators)
Beat odd-even for merging (delay)
Program perfect shuffle machine
Vector sorting algebra
Does removing network constraint help delay?
Exact network lower bound for some N>8
Prove that network lower bound is not O(N logN)
Largest N for which perfect m-sorters exist
Is quicksort optimal minimean restricted uniform?

SEARCHING (Knuth 1972 open problems)

Average case interpolation search
BST deletion (effect of checking null link)
Optimum BST convexity
Hu-Tucker generalize to t-way?
Enumerate AVL trees
AVL trees asymptotically optimal?
Analyze 2-3 trees
Average number of nodes in DSTs
Optimum DSTs
Random probing with tertiary clustering
Linear probing with buckets
Nonrandom hash function for open addressing?
Single hashing vs. random probing
Worst possible single hashing?
Restricted single hashing
Deletion for linear probing
Combinatorial hashing

Solved problems 1972-1997

Avg. length of longest increasing subsequence
Shellsort worst case
Average case for Batcher’s sort
Variance for selection sort quantity
Heapsort average case
Enumerate leftist trees
Minimum average comparisons for merging
Networks with m-sorter modules
Program perfect shuffle machine
Batcher’s has optimal delay
Lower bound for sorting networks
Interpolation search analysis
Optimum BST convexity
AVL tree enumeration
Average node count in DSTs
Optimum DSTs
Random probing with tertiary clustering
Linear probing with buckets
Deletion for linear probing

New problems 1997

- Graph labelling conjecture
- Random sorting networks
- Variance of linear probing
- Associative block designs

Three 30-year open problems

SORTING

- Average case of Shellsort
  - O(N log N) for some increment sequence?
  - Compute best increment sequence?

SEARCHING

- Balanced trees asymptotically optimal?
  - Oscillation in leading term??
  - Simpler top-down algorithm?

NETWORKS

- Practical networks of depth clgN for small c?

N	lgN	GOAL	Batcher
1 thousand	**10**	**20**	**25**
1 million	**20**	**40**	**100**
1 billion	**30**	**60**	**225**