12.排序

（c3）希尔排序：更佳的序列

邓俊辉

deng@tsinghua.edu.cn

d-Sorting an $\mathcal{O}(d)$-Ordered Sequence in $\mathcal{O}(dn)$ Time

❖ If both $\boxed{g}$ and $\boxed{h}$ are in $\boxed{\mathcal{O}(d)}$

we can d-sort the sequence in $\boxed{\mathcal{O}(dn)}$ time ...

❖ 1) re-arrange the sequence as a 2D matrix with $\boxed{d}$ columns

2) each element is swapped with

$$\mathcal{O}(\ (g - 1)*(h - 1)/d\ )\ =\ \mathcal{O}(d)$$

elements

❖ Since this holds for all elements, $\mathcal{O}(dn)$ steps are enough

	[i]	could be smaller than S[i]	must be no less than S[i]

$\longleftarrow$ (g-1)(h-1) $\longrightarrow$

PS Sequence

- Papernov & Stasevic, 1965 //also called Hibbard's sequence

  $\mathcal{H}_{ps} = \boxed{\mathcal{H}_{shell} - 1} = \{ 1, 3, 7, 15, 31, 63, \dots, \boxed{2^k - 1}, \dots \}$

- Different items $\boxed{\text{may not}}$ be relatively prime, e.g. $\boxed{h_{2k} = hk*(hk + 2)}$

  But $\boxed{\text{adjacent}}$ items $\boxed{\text{must}}$ be, since $\boxed{h_{k+1} - 2*hk = \boxed{1}}$

- Shellsort with $\mathcal{H}_{ps}$ needs

  $\boxed{\mathcal{O}(\log n)}$ outer iterations and

  $\boxed{\mathcal{O}(n^{3/2})}$ time to sorts a sequence of length n

- Why ...

PS Sequence

- Let $\boxed{h_t}$ be the h closest to $\boxed{\sqrt{n}}$ //$h_t = \Theta(n^{1/2})$

1) Consider those iterations where $\boxed{k \boxed{>} t}$

   ∵ there would be $\mathcal{O}(\boxed{n/h_k})$ elements in each of the $\boxed{h_k}$ columns

   ∴ we can $\boxed{\text{insertionsort}}$ each column in $\mathcal{O}(\boxed{(n/h_k)^2})$ time

   ∴ each h_k-sorting costs $\mathcal{O}(\boxed{n^2/h_k})$ time

   ∴ all these iterations cost

   $\mathcal{O}(\boxed{2 \times n^2/h_t}) = \mathcal{O}(\boxed{n^{3/2}})$

$k \leq t$

$h_k \leq h_t$

$k > t$

$h_k > h_t$

$k = t$

$h_k = h_t$

PS Sequence

2) Consider those iterations where $\boxed{k \boxed{\le} t}$

∵ $\boxed{h_{k+1}}$ and $\boxed{h_{k+2}}$ are relatively prime and are both in $\mathcal{O}(\boxed{h_k})$

∴ each h_k-sorting costs $\mathcal{O}(\boxed{h_k \times n})$ time

∴ all these iterations cost $\mathcal{O}(\boxed{n \times 2h_t}) = \mathcal{O}(\boxed{n^{3/2}})$ time

❖ This upper bound is $\boxed{\text{TIGHT}}$

❖ How about the average cases?

$\mathcal{O}(\boxed{n^{5/4}})$ based on simulations

but not proved yet

$k \le t$

$h_k \le h_t$

$k > t$

$h_k > h_t$

$k = t$

$h_k = h_t$

Pratt's Sequence

- Pratt, 1971

  $\mathcal{H}_{pratt} = \{\ 1,\ 2,\ 3,\ 4,\ 6,\ 8,\ 9,\ 12,\ 16,\ \ldots,\ \boxed{2^p3^q},\ \ldots\ \}$

- With $\mathcal{H}_{pratt}$,

  Shellsort sorts a sequence of length n in $\mathcal{O}(\boxed{n\log^2 n})$ time //Why?

- And, wait a minute, but ...

  consecutive numbers in $\mathcal{H}_{pratt}$

  are $\boxed{\text{NOT}}$ all relatively prime ...

Pratt's Sequence

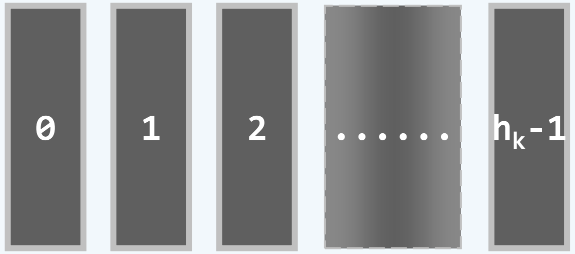
- From $\boxed{(2,\ 3)}$-ordered to $\boxed{1}$-ordered

  $\because$ x(2, 3) = $\boxed{1}$

  $\therefore$ to the right of each element in a $\boxed{(2,\ 3)}$-ordered sequence,

  only the $\boxed{\text{next}}$ element can be smaller

  $\therefore$ it costs $\boxed{\mathcal{O}(n)}$ time to sort such a sequence



- From $\boxed{(2*h_k,\ 3*h_k)}$-ordered to $\boxed{h_k}$-ordered

  divide S into h_k subsequences, each of which is $\boxed{(2,\ 3)}$-ordered

  $\therefore$ it costs altogether $\boxed{\mathcal{O}(n)}$ time to sort them resp.

- $\because$ there are altogether $\boxed{\mathcal{O}(\log^2 n)}$ iterations //why

  $\therefore$ we need time $\boxed{\mathcal{O}(n)} \times \boxed{\mathcal{O}(\log^2 n)} = \boxed{\mathcal{O}(n*\log^2 n)}$

Sedgewick's Sequence

- $\boxed{\text{Pratt}}$'s sequence performs best asymptotically but

  requires too many iterations and hence

  is not good for pre-sorted sequences

- $\boxed{\text{Sedgewick}}$'s sequence: a combination of $\boxed{\text{PS}}$'s sequence with $\boxed{\text{Pratt}}$'s

  { 1, 5, 19, 41, 109, 209, 505, 929, 2161, 3905, 8929, ... }

  of form $\boxed{9 \times 4^k - 9 \times 2^k + 1}$ or $\boxed{4^k - 3 \times 2^k + 1}$

  worst $\mathcal{O}(\boxed{n^{4/3}})$, average $\mathcal{O}(\boxed{n^{7/6}})$,

  best performance in practice

- Is there a step sequence with $\boxed{\mathcal{O}(n\log n)}$ worst-case performance?