12.排序

（b3）选取：通用算法

世兄的才名，弟所素知的。在世兄是数万人里头
选出来最清最雅的，至于弟乃庸庸碌碌一等愚人，
忝附同名，殊觉玷辱了这两个字。

邓俊辉

deng@tsinghua.edu.cn

尝试：蛮力

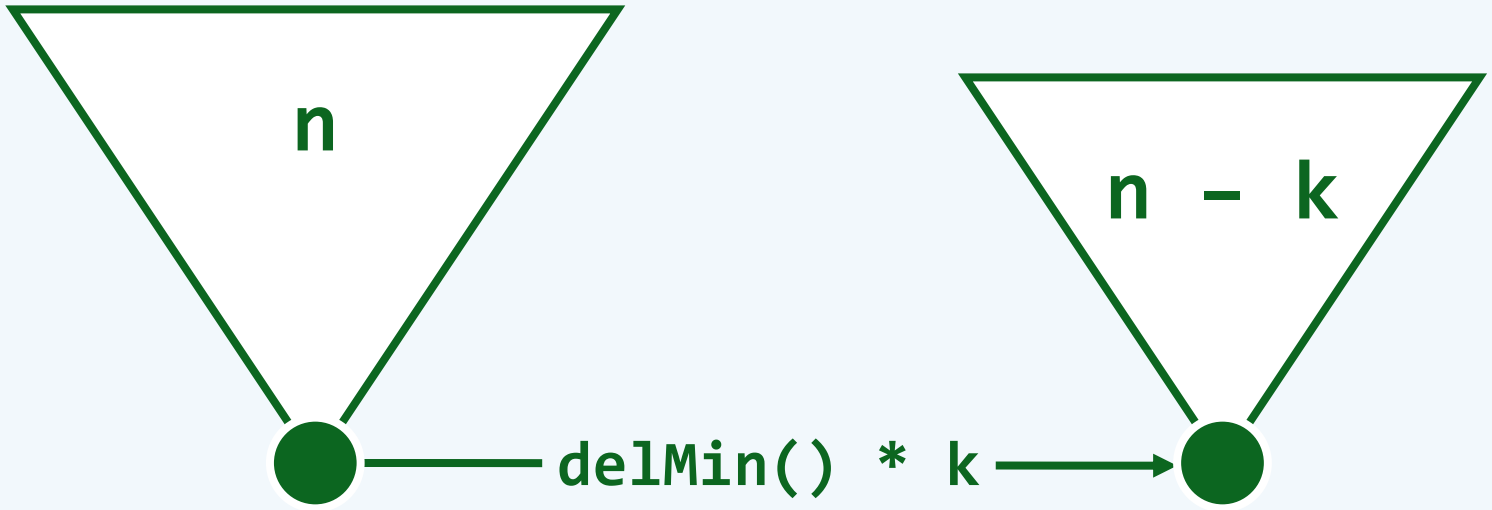
- 对A排序　　//O(nlogn)

从前向后行进k步　　//O(k) = O(n)

尝试：堆（A）

❖ 将所有元素组织为`小顶堆`　　//O(n)

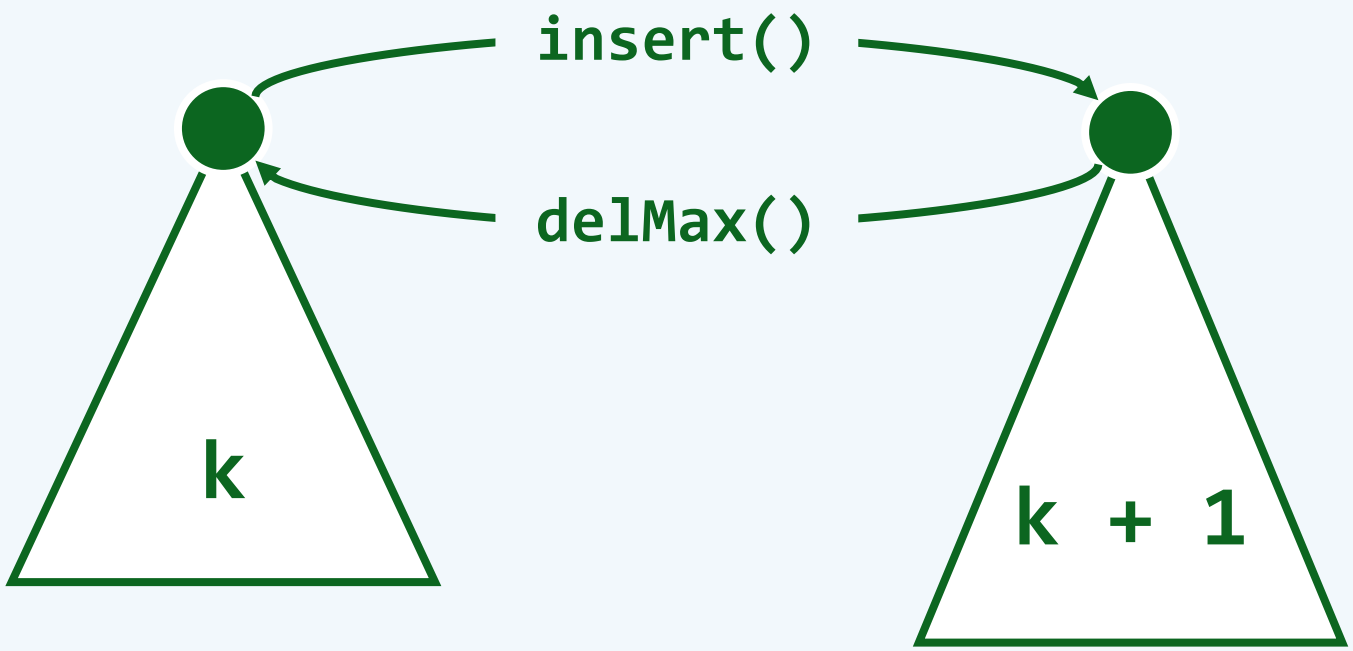
连续调用k次`delMin()`　　//O(klogn)



尝试：堆（B）

❖ 任选（比如前）k个元素，组织为大顶堆 //O(k)

对于剩余的n - k个元素，各调用一次insert()和delMax() //O(2*(n - k)*logk)



尝试：堆（C）

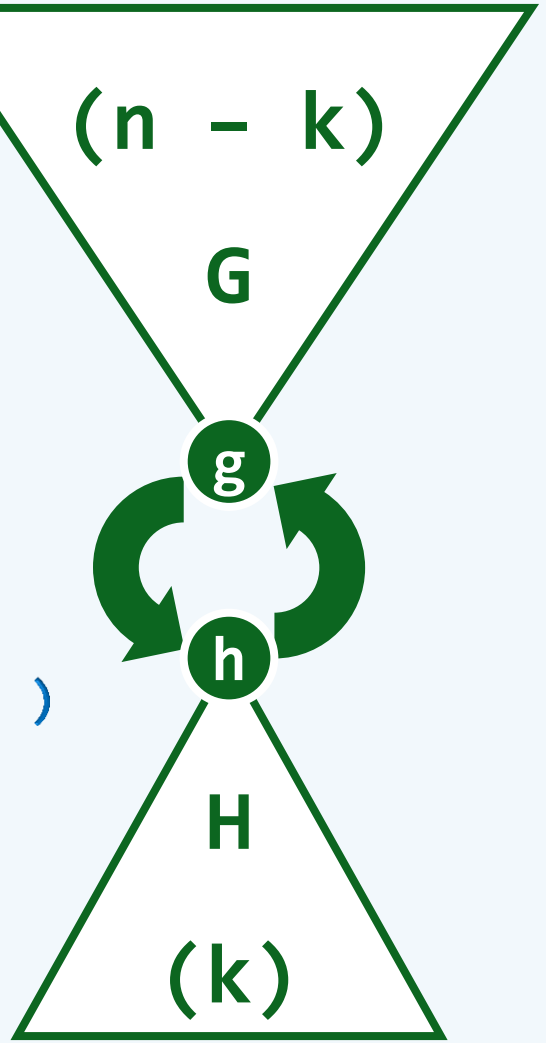
❖ H：任取k个元素，组织为大顶堆 //O(k)

G：其余n - k个元素，组织为小顶堆 //O(n - k)

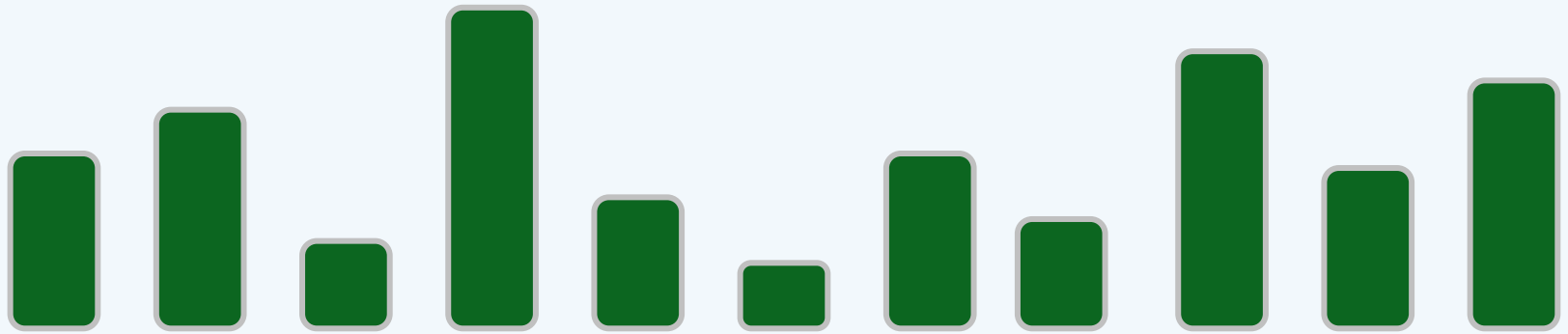
❖ 反复地： 比较h和g //O(1)

如有必要，交换之 //O(2 × (logk + log(n - k)))

直到： h ≤ g //O(min(k, n - k))



尝试：计数排序



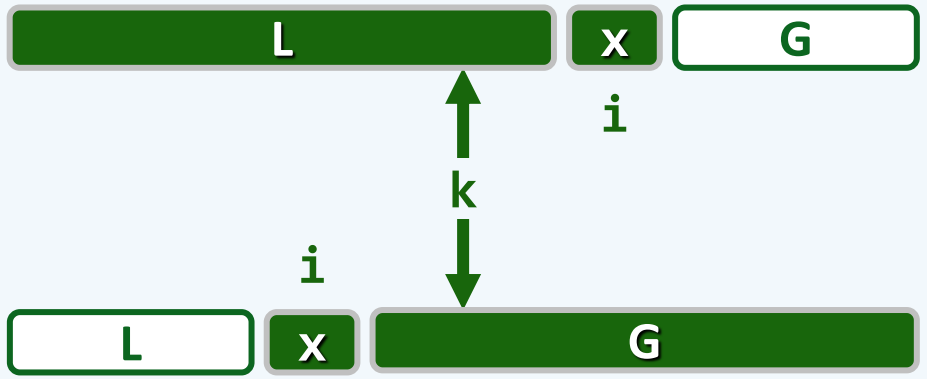
下界与最优

- 是否存在**更快**的算法？
- $\Omega(n)$ ！
- 所谓**第k大**，是相对于序列**整体**而言

  在访问每个元素**至少一次**之前，绝无可能确定
- 反过来，是否存在$O(n)$的算法？

quickSelect()

```
template <typename T> void quickSelect( Vector<T> & A, Rank k ) {
   for ( Rank lo = 0, hi = A.size() - 1; lo < hi;  ) {
      Rank i = lo, j = hi; T pivot = A[lo];
      while ( i < j ) { //O(hi - lo + 1) = O(n)
         while ( i < j && pivot <= A[j] ) j--; A[i] = A[j];
         while ( i < j && A[i] <= pivot ) i++; A[j] = A[i];
      } //assert: i == j
      A[i] = pivot;
      if ( k <= i ) hi = i - 1;
      if ( i <= k ) lo = i + 1;
   } //A[k] is now a pivot
}
```



linearSelect()

Let Q be a small constant

0. if (n = |A| < Q) return trivialSelect(A, k)

1. else divide A evenly into n/Q subsesquences (each of size Q)

2. Sort each subsequence and determine n/Q medians //e.g. by insertionsort

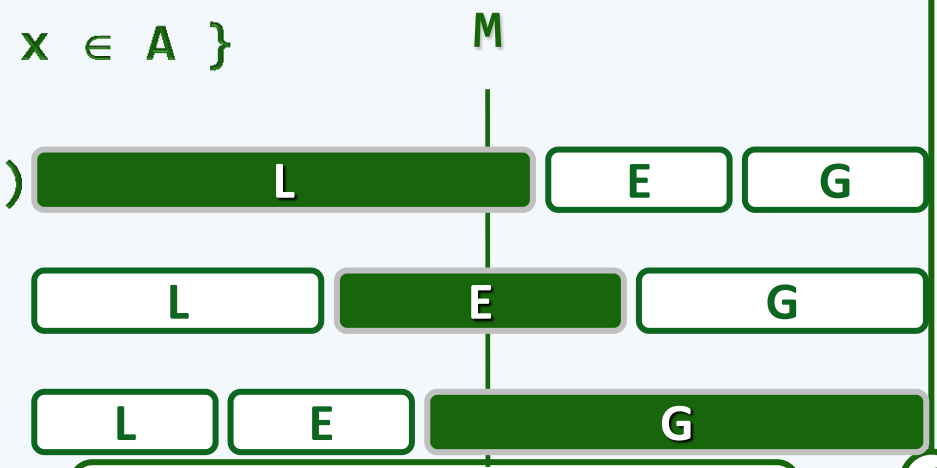
3. Call linearSelect to find M, median of the medians //by recursion

4. Let L / E / G = { x < / = / > M | x ∈ A }

5. if (k ≤ |L|) return linearSelect(L, k)

   if (k ≤ |L| + |E|) return M

   return linearSelect(G, k - |L| - |E|)



复杂度

- 将linearSelect()算法的运行时间记作T(n)
- 第0步：O(1) = O(QlogQ) //递归基：序列长度|A| ≤ Q
- 第1步：O(n) //子序列划分
- 第2步：O(n) = O(1) × n/Q //子序列各自排序，并找到中位数
- 第3步：T(n/Q) //从n/Q个中位数中，递归地找到全局中位数
- 第4步：O(n) //划分子集L/E/G，并分别计数 —— 一趟扫描足矣
- 第5步：T(3n/4) //为什么...

复杂度

- 在某种意义上，如上所确定的M必然不偏不倚

  至少各有n/4个元素，不小于/不大于M

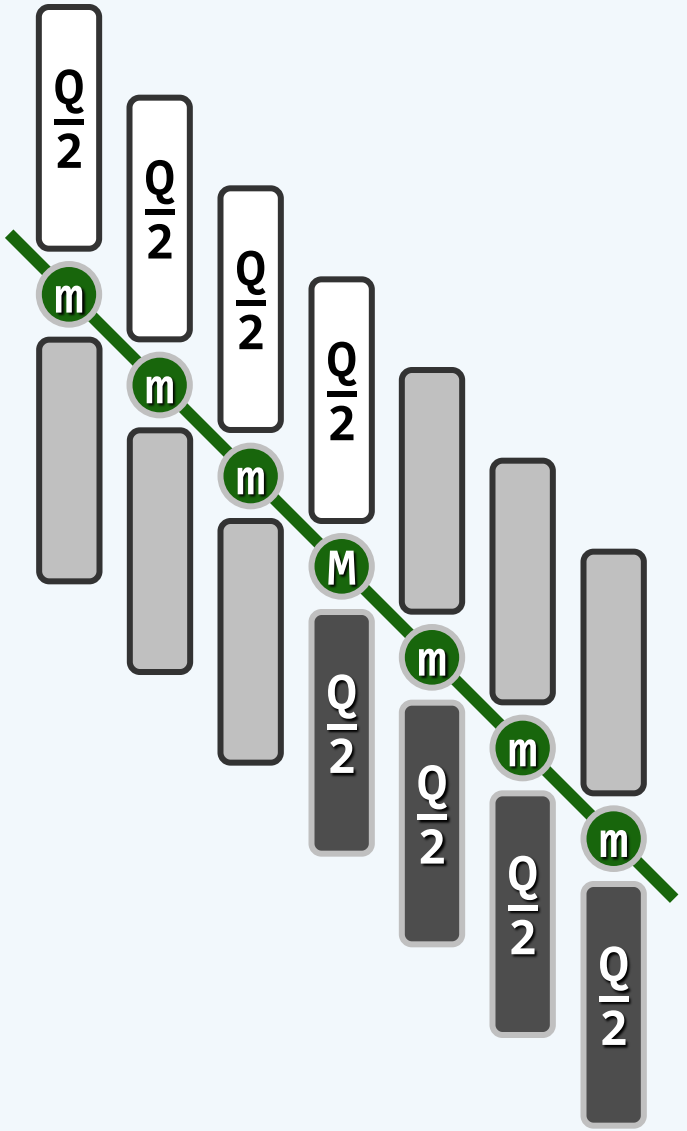
- n/Q个中位数中，至少半数不小于M

  而它们又各自不大于至少Q/2个元素

  $$n/Q \;/\; 2 \;*\; Q/2 \;=\; n/4$$

- $\min(\,|L|,\,|G|\,) + |E| \;\geq\; n/4$

  $\max(\,|L|,\,|G|\,) \;\leq\; 3n/4$



复杂度

- $T(n) = O(n) + T(n/Q) + T(3n/4)$
- 为使之解作线性函数，只需保证

$$n/Q + 3n/4 < n$$

或等价地

$$1/Q + 3/4 < 1$$

- 比如，若取 $Q = 5$，则存在常数c，使得

$$T(n) = cn + T(n/5) + T(3n/4)$$

$$T(n) = O(20cn) = O(n)$$