2021 年算法设计与分析论文考试往年题整理

I. Mathematical Analysis of Algorithms

1. 什么是in situ permutation？它对于存储空间的要求是什么？

答：输入是一维数组 $(x_1,x_2,\cdots,x_n)$，以及 $\{1,2,\cdots,n\}$ 上的一个置换 p，输出是 $(x_{p(1)},x_{p(2)},\cdots,x_{p(n)})$。除了输入以外，只用 $O(1)$ 辅助存储空间。对于每个 k，可以计算出 $p(k)$，但是不能再给 $p(k)$ 赋新的值。

2. 语句“k:=$p(k)$”的执行次数记为 $a(\pi)$，那么 $a(\pi)$ 的最大值、最小值、平均值各是多少？

答：$a(\pi)$ 的最大值是 $\frac{n(n-1)}{2}$，最小值是0，平均值是 $O(n\ln n)$。

3. 在增加了tally变量后，第6行代码循环次数平均减少多少次？（注：论文中的结论有误）

答：平均减少了

$$\sum_{2\le i<j\le n}\frac{1}{(j-i+2)(j-i+1)}=\sum_{3\le r\le n}\frac{n+1-r}{r(r-1)}$$
$$=(n+1)\sum_{3\le r\le n}\frac{1}{r(r-1)}-\sum_{3\le r\le n}\frac{1}{r-1}$$
$$=(n+1)\left(\frac{1}{2}-\frac{1}{n}\right)-\sum_{2\le r\le n-1}\frac{1}{r}$$
$$=(n+1)\left(\frac{1}{2}-\frac{1}{n}\right)-(H_{n-1}-1)$$
$$=\frac{1}{2}(n+1)-H_n.$$

4. 对于原地重排列算法，如果能够在每个数据上增加一个比特用于辅助算法计算，则算法第六行在改进后的平均执行次数可能会是多少？请简要说明改进的方法和分析的依据。

答：可以通过这多出来的一个比特，把遍历过的位置都标记下，避免重复遍历寻找圈头的位置，这样总运行时间就优化为 $O(n)$。

5. 论文中的In Situ Permutation算法最坏情形的时间复杂度是多少？对应的输入是怎样的？

答：最坏情况发生在输入为 $(p(1),p(2),\cdots,p(n))=(2,3,\cdots,n,1)$ 时，对应的时间复杂度为 $O(n^2)$。

6. 如果按论文中所述增加了 tally 变量，其最坏情形的时间复杂度是怎样的？对应的输入又是怎样的？

答：最坏情况发生在输入为 $(p(1),p(2),\cdots,p(n))=(2,3,\cdots,n,1)$ 时，对应的时间复杂度为 $O(n^2)$。

II. An Analysis of Alpha-Beta Pruning

1. 什么是 alpha-beta pruning？

答：alpha-beta 剪枝是对 game tree 求值的方法，它用两个值 alpha（下界）和 beta（上界）来帮助在深度优先搜索中进行剪枝。

2. 什么是 game tree？

答：game tree 是表示双人博弈的根树，每个节点对应于一个局面，每个分支对应

一种走法选择，根节点是第 0 层，单数层和双数层分别对应于两个选手可能的走法导致的局面，树叶是终局。每个节点有一个值，终值的值直接给定，非终局的值按照最大规则或最小规则确定，最大规则规定父节点的值等于子女节点的值，最小规则规定父节点的值等于子女节点的最小值，两个选手分别应用最大规则和最小规则，根节点的值称为游戏的值，代表双方最佳走法的结果。

3. 什么是 critical position？

答：对于第 L 层的节点，用序列 $a_1a_2\cdots a_L$ 表示它的位置，a_i 表示在第 i 层选择了第 a_i 个分支，单数位全是 1 或双数位全是 1 的序列所对应的节点称为 critical positions。

4. 什么是 deep cutoff？它对于 alpha-beta pruning 有什么影响？

答：deep cutoff 就是指引入 alpha 下界之后，带来的剪枝效果可以到达兄弟节点的子女节点。

对于好的节点排序，deep cutoff 对于 alpha-beta 剪枝只有次一阶的影响；但是对于不好的排序，则有更大的影响。

5. 什么是 dewey decimal system？

答：dewey decimal system 是用一组长度为 L 的正整数序列表示 game tree 中的在第 L 层的任意一个 position 的方法。具体来说，根节点对应空序列，而其他节点，若其父亲节点的序列是 s，且自己是父亲节点的第 d 个儿子，则自己的序列为 s，d。

6. 什么是 perfect ordering assumption？

答：perfect ordering assumption 是指每个非叶子节点的第一个儿子是最好的选择（返回最优的值）。

7. 在什么意义下，alpha-beta pruning 是最优的？

答：对于任何一个 game tree 和任何一个计算它的根节点值的算法，总是可以通过一个重排（改变兄弟节点的顺序），使得在此重排下，凡是 alpha-beta 剪枝方法检查过的终局，也要被该算法检查；而且在此重排下，当根节点的值不是正负无穷大时时，alpha-beta 剪枝方法恰好检查的是那些 critical positions。

III. Smoothed Analysis: An Attempt to Explain the Behavior of Algorithms in Practice

1. 作者提出平滑分析，是为了解释什么现象？

答：The goal of smoothed analysis is to explain why some algorithms have much better performance in practice than predicted by the traditional worst-case analysis.

2. 平滑分析中为什么要引入扰动？

答：Practical data is often subject to some small degree of random noise.

3. 在平滑分析中，σ 的作用是什么？

答：The smoothed complexity of an algorithm measures the performance of the algorithm both in terms of the input size n and in terms of the magnitude sigma of the perturbation. By varying sigma between zero and infinity, one can use smoothed analysis to interpolate between worst-case and average-case analysis. The dependence on the magnitude sigma is essential and much of the work in smoothed analysis demonstrates that noise often makes a problem easier to solve.

4. 平滑分析有什么局限？如何加强？

答：Smoothed analysis cannot always produce the best explanations for all problems, since it is difficult to capture the essential aspects of practical input instances.
One way to strengthen the smoothed analysis framework is to improve the model of the

formation of input instances. A simpler way to strengthen smoothed analysis is to restrict the family of perturbations considered.

5. 举例说明平滑分析有什么实际用处。

答：Smoothed analysis and input perturbation can help to solve some problems more efficiently. For example, the smoothed analysis of Gaussian elimination in this paper suggests a more stable solver for the linear system $A\boldsymbol{x} = \boldsymbol{b}$.

6. 在引入“平滑分析”之前，传统的“最坏情况分析”和“平均情况分析”遇到了什么问题？

答：There is a big gap between the theoretical analysis and practical performance for lots of algorithms. People will see many algorithms work well in practice, in spite of having a poor, sometimes exponential, worst-case running time. To explain this inconsistency, people employ average-case analysis and measure the expected running time of an algorithm on some distribution of inputs. However, since it is rare that one can determine or cleanly express these distributions, and the distributions can vary greatly between one application and another, average-case analyses usually employ distributions with concise mathematical descriptions, such as Gaussian random vectors. The drawback of using such distributions is that the inputs actually encountered in practice may bear very little resemblance to the inputs that are likely to be generated by such distributions.

7. 什么是平滑复杂度？它有何意义？又有何局限？

答：In smoothed analysis, we assume that an input to an algorithm is subject to a slight random perturbation. The smoothed measure of an algorithm on an input instance is its expected performance over the perturbations of that instance. Thus the smoothed complexity of an algorithm is defined to be the maximum smoothed measure over input instances.

The smoothed complexity of an algorithm measures the performance of the algorithm both in terms of the input size n and in terms of the magnitude σ of the perturbation. By varying σ between zero and infinity, one can use smoothed analysis to interpolate between worst-case and average-case analysis.

The smoothed analysis and smoothed complexity cannot explain all the phenomena. For many problems, there may be better explanations.

8. 进行算法的平均情况复杂性分析时，需假定输入实例的概率分布。通常为了方便计算，我们会假定输入实例的分布是完全随机的，这么做会有何问题？为什么？

答：我们假定输入实例的分布是完全随机的，但这与实际情况不相符（就像random graphs和random matrices与实际不符一样），这会严重干扰其后的分析（论文中说“Random objects have special properties with exponentially high probability, and these special properties might dominate the average-case analysis”）。

之所以出现这种不相符，是因为输入实例的多个维度之间是相互作用的，而我们并不知道这种作用关系，只能简单地假设各个维度之间的独立性从而“完全随机”地生成实例。事实上，这种random objects很可能是very special objects（论文中提到“We argue that “random matrices” are very special matrices”）。

9. 中心极限定理（central limit theorem）告诉我们：一个随机事件的出现受到许多相互独立的随机因素的影响，如果每个因素所产生的影响都很微小时，总体的影响可以看作是服从正态分布的。请问这一定理与论文中的算法平滑分析有什么关系？

答：论文中说“practical data is often subject to some small degree of random noise”，

根据中心极限定理，我们可以认为输入实例符合正态分布（高斯分布），因此论文中说“the family of Gaussian distributions provides a natural model of noise or perturbation”，然后很自然地以高斯扰动模型为例来定义和分析平滑复杂度（smoothed complexity）。

IV. Primes is in P

1. 请说明文章标题“PRIMES is in P”中的“PRIMES”指什么？

答：PRIMES is the decisional problem of determining whether or not a given integer n is prime.

2. 文章摘要：“We present an unconditional deterministic polynomial-time algorithm determines whether an input number is prime or composite.”请分别简单解释“unconditional”，“deterministic”，“polynomial-time”这三个词的含义。

答：“unconditional”指 AKS 算法不依赖于一些未证的假设；

“deterministic”指对于同一个输入，AKS 算法总是返回确定的输出（素数或者合数），不像其他随机算法有时输出素数有时输出合数；

“polynomial-time”是算法时间复杂度是输入数 n 的位数 $\log n$ 的多项式函数。