Practice Questions for the Final for CS 574: Spring 2014

Sariel Har-Peled*

May 10, 2014

"Then you must begin a reading program immediately so that you man understand the crises of our age," Ignatius said solemnly. "Begin with the late Romans, including Boethius, of course. Then you should dip rather extensively into early Medieval. You may skip the Renaissance and the Enlightenment. That is mostly dangerous propaganda. Now, that I think about of it, you had better skip the Romantics and the Victorians, too. For the contemporary period, you should study some selected comic books."

"You're fantastic."

"I recommend Batman especially, for he tends to transcend the abysmal society in which he's found himself. His morality is rather rigid, also. I rather respect Batman."
– John Kennedy Toole, A confederacy of Dunces.

Version 1.02

The following is a list of problems that might be included in the final. While some of these problems are too hard for a final exam, I might take some subset of a problem as a question in the final exam. The final would contain at least one problem that is not on this list, and at least two problems from this list.

(Yes, there are typos - I would fix them over the next few days.)

1. Problems

1. **QuickSort**. (25 PTS.)
Using linearity of expectation, prove that the running time of **QuickSort** is $O(n \log n)$.

2. Smallest disk containing 5 points (25 PTS.)
Given a set P of n points in the plane, describe an algorithm that in expected linear time computes the smallest disk containing 5 points of P. You can assume that given 5 points one can compute the smallest disk containing them in constant time.
Hint: Think closest pair algorithm.

3. Balanced pink path. (25 PTS.)
Consider a balanced tree T of height h with every node having three children, and consider coloring its edges by three colors, say, red, blue and (of course) pink. Give an upper-bound and a lower-bound the probability that there is a pink path from the root of T to one of its leafs that is all made out of pink edges.

*Department of Computer Science; University of Illinois; 201 N. Goodwin Avenue; Urbana, IL, 61801, USA; sariel@illinois.edu; http://sarielhp.org/. Work on this paper was partially supported by a NSF AF award CCF-0915984.

4. Coupon collecting with exchanges. (25 PTS.)

There is a set C of n different coupons, at every iteration you pick a random coupon. The purpose of the game is to collect all possible coupons. The twist here is that if you are missing i coupons, you can exchange one of these missing coupons for $f(i)$ of your extra coupons (after all, you are going to have a lot of extra coupons that are identical - so why not get rid of them?). Describe a strategy as efficient as possible, that takes the (expected) minimum number of rounds to collect all coupons. Prove your answer. do this for the following two cases:

(A) $f(i) = \lceil n/i^2 \rceil$.

(B) $f(i) = \lceil n/i \rceil$ (in this case, prove a lower bound showing that no better strategy exists).

5. Stream select. (25 PTS.)

You are given a stream of random numbers $x_1, \ldots, x_n$. The numbers are selected independently from the same distribution. You do *not* know the distribution, but you do know n in advance. (You can assume all the numbers in the stream are distinct.)

Here you are in a streaming situation – you can read the data, but you do not have the space to store it. So, if the algorithm decides to throw away a number, and it needs it later then it had failed.

(A) Assume you are given only $O(m)$ space, and a parameter k. Describe an algorithm with high probability outputs the element of rank k in the input, and uses only m space. Provide a lower bound and upper bound on the minimum m for which your algorithm would work. For credit, m has to be smaller than, say, $n^{0.9}$. For full credit you need to prove tight upper and lower bounds.

(B) You are allowed to consult with an oracle t times, where given a element x_i, the oracle tells you the rank of x_i in the input. Design an efficient algorithm that uses as little space m as possible, and the oracle, and outputs the element of rank k in the stream.

What is (asymptotically) the minimum value of t, such that using $m = O(1)$ space, your algorithm outputs the element of rank k, and the algorithm succeeds with high probability? Give a strong argument (or even better, prove) that your value of t is optimal.

6. Unconditional fast **QuickSort**. (25 PTS.)

Using only conditional expectations (no Chernoff, Martingales or similar tools), prove that **QuickSort** takes $O(n \log n)$ time with probability $\geq 1 - 1/n^5$.

7. Chernoff inequality. (25 PTS.)

Let $X_1, \ldots, X_n$ be n independent random variables such that $\mathbf{Pr}[X_i = 0] = \mathbf{Pr}[X_i = 1] = 1/2$. Let $Y = \sum_{i=1}^{n} X_i$, and $\mu = \mathbf{E}[X_i]$.

Prove that there are positive constants c_1, c_2, c_3, independent of n, such that if $\delta < c_1$, then $\mathbf{Pr}[(1-\delta)\mu \leq Y \leq (1+\delta)\mu] \leq c_2 \exp(-c_3' \delta^2 \mu)$.

8. Azuma's inequality. (25 PTS.)

Prove Azuma's inequality (as in the previous question, the exam might ask you to do some portion of the proof, not necessarily the whole thing from scratch.

9. Good coloring. (25 PTS.)

Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph with n vertices and m edges. Prove that there is always a coloring of the graph with k colors, such that the number of edges that are monochromatic (i.e., both endpoints have the same color), is at most m/k. Prove that this is tight in the worst case (i.e., one can not do better).

10. Expander. (25 PTS.)

Consider a graph G over n vertices (assume n is even) that is the union of t matchings, where t is a sufficiently large constant. Prove that G is a good expander (specify exactly the parameters for which your graph is an expander – the smaller the t, the better it is.

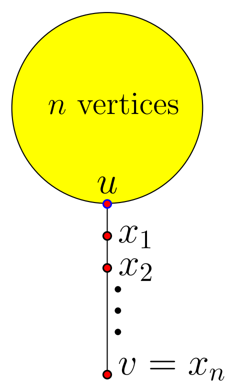
11. Deterministic discrepancy. (25 PTS.)
Given a set U with n elements, and a family $\mathcal{F}$ of m subsets of U, describe an algorithm that computes a coloring of U by $-1, +1$, such that discrepancy of every set of $\mathcal{F}$ is at most $O(\sqrt{n \log m})$.

12. k-sets. (25 PTS.)
Given a set P of n points in the plane and a parameter k, prove that the number of distinct k-sets is at most $O(nk)$.
A subset S of P is a ***k-set*** if there is a halfplane h^+, such that $|S| \leq k$, and $S = P \cap h^+$.

13. Random walk on the grid. (25 PTS.)
Prove that a random walk starting from the origin in the two dimensional grid revisits the origin infinite number of times in expectation.

14. 2SAT. (25 PTS.)
You are given a formula F with n variables and m clauses. Describe an algorithm that solves 2SAT, and prove that its running time is polynomial (in expectation if it is randomized).

15. Lollipop. (25 PTS.)
Consider the lollipop :



It top part is K_n, and its bottom is a path of length n. Provide upper and lower bounds on the following quantities (as tight as possible). Prove your answer:

(A) The expected time of the random walk starting at u to arrive at v.
(B) The expected time of the random walk starting at v to arrive at u.
(C) The expected time of the random walk starting at v to visit all the vertices in the graph.
(D) The expected time of the random walk starting at u to visit all the vertices in the graph.

16. Prove measure concentration on the sphere. (25 PTS.)
Prove the following.

Theorem 1.1 (Measure concentration on the sphere.). *Let $A \subseteq \mathbb{S}^{(n-1)}$ be a measurable set with $\mathbf{Pr}[A] \geq 1/2$, and let A_t denote the set of points of $\mathbb{S}^{(n-1)}$ in distance at most t from A, where $t \leq 2$. Then $1 - \mathbf{Pr}[A_t] \leq 2\exp(-nt^2/2)$.*

Similarly, prove Levy's lemma.

17. JL Lemma works for angles. (25 PTS.)

Show that the Johnson-Lindenstrauss lemma also $(1 \pm \varepsilon)$-preserves angles among triples of points of P (you might need to increase the target dimension however by a constant factor).
Hint: For every angle, construct a equilateral triangle that its edges are being preserved by the projection (add the vertices of those triangles [conceptually] to the point set being embedded). Argue, that this implies that the angle is being preserved.

18. VC dimension. (25 PTS.)
Provide a bound, as tight as possible, on the VC dimension of the range space, where the ground set is the plane, and the ranges are triangles.Flip and flop (25 PTS.)

(A) Let $b_1, \ldots, b_{2m}$ be m binary bits. Let Ψ be the set of all permutations of $1, \ldots, 2m$, such that for any $\sigma \in \Psi$, we have $\sigma(i) = i$ or $\sigma(i) = m + i$, for $1 \le i \le m$, and similarly, $\sigma(m+i) = i$ or $\sigma(m+i) = m+i$. Namely, $\sigma \in \Psi$ either leaves the pair $i, i+m$ in their positions or it exchanges them, for $1 \le i \le m$. As such $|\Psi| = 2^m$.
Prove that for a random $\sigma \in \Psi$, we have

$$\mathbf{Pr}\left[\left|\frac{\sum_{i=1}^m b_{\sigma(i)}}{m} - \frac{\sum_{i=1}^m b_{\sigma(i+m)}}{m}\right| \ge \varepsilon\right] \le 2e^{-\varepsilon^2 m/2}.$$

(B) Let Ψ' be the set of all permutations of $1, \ldots, 2m$. Prove that for a random $\sigma \in \Psi'$, we have

$$\mathbf{Pr}\left[\left|\frac{\sum_{i=1}^m b_{\sigma(i)}}{m} - \frac{\sum_{i=1}^m b_{\sigma(i+m)}}{m}\right| \ge \varepsilon\right] \le 2e^{-C\varepsilon^2 m/2},$$

where C is an appropriate constant. [**Hint:** Use (18A), but be careful.]

(C) Prove the ϵ sample theorem using (18B). This is theorem 20.3.4 in the class notes.

Theorem 1.2 (ε-sample theorem, [VC71]). *There is a positive constant c such that if $(\mathsf{X}, \mathcal{R})$ is any range space with* VC *dimension at most δ, $\mathsf{x} \subseteq \mathsf{X}$ is a finite subset and $\varepsilon, \varphi > 0$, then a random subset $C \subseteq \mathsf{x}$ of cardinality*

$$s = \frac{c}{\varepsilon^2}\left(\delta \log \frac{\delta}{\varepsilon} + \log \frac{1}{\varphi}\right)$$

is an ε-sample for x with probability at least $1 - \varphi$.

19. Convex hulls incrementally (25 PTS.)
Let P be a set of n points in the plane.

(A) Describe a randomized incremental algorithm for computing the convex hull $\mathcal{CH}(\mathsf{P})$. Bound the expected running time of your algorithm.

(B) Assume that for any subset of P, its convex hull has complexity t (i.e., the convex hull of the subset has t edges). What is the expected running time of your algorithm in this case? If your algorithm is not faster for this case (for example, think about the case where $t = O(\log n)$), describe a variant of your algorithm which is faster for this case.

20. Number of minima points. (25 PTS.)
Let Q be a set of n students. Given any subset $X \subseteq Q$ of the students, the university might choose (in an arbitrary fashion) up to k of the students of X to be the representatives of the set X, and let $\ell(X)$ be this set of representatives.

(A) Let $P = \langle p_1, \ldots, p_n \rangle$ be a random permutation of Q. The point p_i is a ***winner*** if it is a representative of $P_i = \langle p_1, \ldots, p_i \rangle$. Prove, that with high probability there are at most $O(k \log n)$ winners.

(B) Let Q be a set of n points, picked uniformly in the unit square $[0,1]^2$. A point $p = (x, y) \in Q$ is a ***minima*** if there is no point $(x', y') \in Q$ such that $x' \leq x$ and $y' \leq y$. Prove, using (A), that with high probability, the number of minima points of Q is $O(\log n)$.

(C) Let Q be a set of n points, picked uniformly in the unit cube $[0,1]^3$. A point $p = (x, y, z) \in Q$ is a ***minima*** if there is no point $(x', y', z') \in Q$ such that $x' \leq x$, $y' \leq y$ and $z' \leq z$. Prove, using (A) and (B), that with high probability, the number of minima points of Q is $O(\log^2 n)$.

References

[VC71] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16:264–280, 1971.