KTH Computer Science and Communication

Communication Complexity: Problem Set 3

Due: October 21, 2012. Submit as a PDF file by e-mail to `lauria at kth dot se` with the subject line `Problem set 3: ⟨your name⟩`. Name the PDF file `PS3_⟨YourName⟩.pdf` (with your name coded in ASCII without national characters). Solutions should be written in LaTeX or some other math-aware typesetting system. Please try to be precise and to the point in your solutions and refrain from vague statements. In addition to what is stated below, the general rules stated on the course webpage always apply.
Collaboration: Discussions of ideas in groups of two to three people are allowed, but you should write down your own solution individually and understand all aspects of it fully. For each problem, state at the beginning of your solution with whom you have been collaborating.
Reference material: For some of the problems, it might be easy to find solutions on the Internet, in textbooks or in research papers. It is not allowed to use such material in any way unless explicitly stated otherwise. You can refer without proof to anything said during the lectures on in the lecture notes, except in the obvious case when you are specifically asked to show something that we claimed without proof in class. It is hard to pin down 100% formal rules on what all this means—when in doubt, ask the lecturer.
About the problems: Some of the problems in the problem sets are meant to be quite challenging and you are not necessarily expected to solve all of them. A total score of around 70 points should be enough for grade E, 125 points for grade C, and 180 points for grade A on this problem set. Any corrections or clarifications will be posted on the course webpage `www.csc.kth.se/utbildning/kth/kurser/DD2441/semteo12/`.

1 (10 p) Recall that for a function $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$, where we can view the inputs x and y as integers in $[0, 2^n - 1]$, we let M_f denote the $n \times n$-matrix with (i,j)-entry $M_{i,j} = f(i,j)$. We define the *rank* of f as $\mathrm{rank}(f) = \mathrm{rank}_{\mathbb{R}}(M_f)$, i.e., the rank of the matrix M_f computed over the field of reals. Prove that the deterministic two-party communication complexity of f is upper-bounded by $D(f) \leq \mathrm{rank}(f) + 1$.

2 (10 p) A *decision tree* T over variables $x_1, \ldots, x_n$ is a binary tree such that every internal vertex is labelled by a variable x_i and the two edges to its left and right child are labelled 0 and 1, respectively, and such that all leaves are labelled by 0 or 1. T defines a function $f_T : \{0,1\}^n \mapsto \{0,1\}$ in the natural way by letting $f_T(\alpha)$ be the value of the leaf reached when walking from the root of T along edges according to α. More formally, given an assignment $\alpha = (\alpha_1, \ldots, \alpha_n)$ to the variables $x_1, \ldots, x_n$, we start at the root and at each internal vertex v, which is labelled by x_i, say, we follow the edge to the child of v whose edge is labelled by α_i, until we reach some leaf, the label of which is the value of $f_T(\alpha)$. T is a *decision tree for* f if $f_T = f$. The *depth* of a tree T is the length of a longest path in T from the root to some leaf, and the *decision tree complexity* $dtc(f)$ of f is the smallest depth of any decision tree for f.

For arbitrary functions $f : \{0,1\}^n \mapsto \{0,1\}$ and $g : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$, let the composed function $F : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ be defined by $F(x,y) = f(g(x_1,y_1), g(x_2,y_2), \ldots, g(x_n,y_n))$. Prove that the deterministic two-party communication complexity $D(F)$ is upper-bounded by $D(F) = \mathrm{O}\big(dtc(f) \cdot D(g)\big)$.

3 (10 p) Let P and Q be arbitrary probability distributions over some common finite domain $\Omega = \{\omega_1, \ldots, \omega_n\}$. Recall that the *total variation distance* $V(P,Q)$ of P and Q is defined as $V(P,Q) = \frac{1}{2}\sum_{\omega\in\Omega}|P(\omega) - Q(\omega)|$.

3a Prove that (as claimed in class) an alternative definition of total variation distance is $V(P,Q) = \max_{\Omega'\subseteq\Omega'}\{P(\Omega') - Q(\Omega')\}$ (where for a subset $\Omega' \subseteq \Omega$ we use the short-hand $P(\Omega') = \sum_{\omega\in\Omega'} P(\omega)$).

3b Prove that total variation distance is a metric on probability distributions — i.e., it is symmetric, non-negative, non-zero unless $P = Q$, and satisfies the triangle inequality $V(P_1,P_2) \leq V(P_1,Q) + V(Q,P_2)$ — and that it always holds that $0 \leq V(P,Q) \leq 1$.

4 (10 p) Let P and Q be arbitrary distributions over $\Omega = \{\omega_1, \ldots, \omega_n\}$. Let $\|x\|$ denote the usual Euclidean norm $\|x\| = \|(x_1, \ldots, x_n)\| = \sqrt{\sum_{i=1}^n x_i^2}$. Prove that the Hellinger distance $h(P,Q)$ is the norm of the vector with ith coordinate equal to the difference of the square roots of the probabilities of seeing ω_i according to P and Q, normalized by dividing by $\sqrt{2}$. That is, prove

$$h(P,Q) = \left\|\left(\sqrt{P(\omega_1)/2} - \sqrt{Q(\omega_1)/2}, \ldots, \sqrt{P(\omega_n)/2} - \sqrt{Q(\omega_n)/2}\right)\right\| \ . \tag{1}$$

Then use the equality (1) to argue that Hellinger distance is in fact (as claimed in class) a metric on probability distributions — i.e., it is symmetric, non-negative, non-zero unless $P = Q$, and satisfies the triangle inequality — and that $0 \leq h(P,Q) \leq 1$.

5 (20 p) It is known that total variation distance is lower-bounded by Hellinger distance by $V(P,Q) \leq h(P,Q)\sqrt{2 - h^2(P,Q)}$ for any P and Q. In [BJKS04], this was combined with the inequality $V(P,Q) \geq 1 - 2\delta$ (for particular P and Q) to yield the desired lower bound $h(P,Q) \geq \sqrt{1 - 2\sqrt{\delta}}$. When doing the calculations on a late-night flight home from a week of hard work in Rome, the main lecturer for some reason instead got the (slightly better) bound $h(P,Q) \geq \sqrt{1 - \sqrt{2\delta}}$. Is this correct? Derive the best bound you can to answer this question!

6 (30 p) As Troy mentioned in his guest lectures, the (non-generalized) discrepancy method (covered in lecture 3) does not work very well for set disjointness. Show that although we know that $R(\mathsf{DISJ}_n) = \Theta(n)$, using discrepancy we can never get a better lower bound than $\mathrm{O}(\log n)$.

7 (40 p) Prove or disprove each of the claims below, where X, Y, Z are always assumed to be arbitrary random variables over finite domains and x, y, z arbitrary outcomes of these random variables. For full credit, you should provide for each subproblem (a) a formal proof of the claim if it is true, or if the claim is false a concrete counter-example with a proof that this is indeed a counter-example; (b) an intuitive, brief explanation why the claim is true or false. For partial credit, you can provide either the formal proof or the intuitive explanation, where a compelling informal argument will give higher scores than formal symbol manipulation.

For instance, suppose that the claim is "the entropy $H(X)$ is maximized for random variables X that are uniformly distributed over their domain." Then this claim is true, and a formal proof could go via Kullback-Leibler divergence as we did in class. The informal explanation could be that "$H(X)$ measures the 'uncertainty' of X, and for a fixed domain size this uncertainty is maximized when X is equally likely to be any element in the domain."

7a $H(X \mid Y) = H(Y \mid X)$.

7b $I(X;Y \mid Z) = I(Y;X \mid Z)$.

7c $I(X;Y) \leq H(X)$.

7d $I(X;Y) \leq I(X;Y \mid Z)$.

7e $H(X \mid Y = y) \leq H(X)$.

8 (30 p) When we did the proof of the linear randomized communication complexity of set disjointness, we wanted to lower-bound the mutual information $I(X,Y;\Pi(X,Y))$ of the inputs (X,Y) and the protocol Π run on these inputs with respect to some well-chosen probability distribution μ. However, we instead obtained a bound on $I(X,Y;\Pi(X,Y) \mid D)$ for a distribution ζ on $((X,Y),D)$ such that (X,Y) was a product distribution conditioned on D and the marginal distribution on (X,Y) agreed with μ. We then proved that the inequality

$$I(X,Y;\Pi(X,Y)) \geq I(X,Y;\Pi(X,Y) \mid D) \tag{2}$$

holds in this setting.

A natural question is how much slack there is in (2), or, expressed differently, how much we lose when going from the left-hand side to the right-hand side of the inequality. In order to prove a strong lower bound, clearly we do not want this loss to be too large. Prove that happily, the difference between the left- and right-hand sides in (2) can never be more than the entropy $H(D)$ of the random variable on which we are conditioning.

9 (40 p) Use the discrepancy method (as described in lecture 3) to prove that as n goes to infinity, 99.9% of all functions $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ are super-hard in the sense that they have two-party randomized communication complexity $R(f) = \Theta(n)$.

10 (40 p) Let STRICT-MAJ-XOR be the function that takes two n-bit strings x and y and evaluates to true if a strict majority of the bitwise exclusive ors evaluate to true. Formally,

$$\text{STRICT-MAJ-XOR}(x,y) = \begin{cases} 1 & \text{if } |\{i \in [n] : x_i \oplus y_i = 1\}| > n/2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Try to figure out whether STRICT-MAJ-XOR is easy or hard in the two-party randomized public-coin communication model, and then prove the best upper bound you can (if you think the function is easy) or the best lower bound you can (if it is hard). For a full score, you should get an optimal bound (up to constant factors hidden in the asymptotic notation), but you do not have to prove that the bound you get is in fact optimal.