KTH Computer Science and Communication

Communication Complexity: Problem Set 4

Due: November 25, 2012. Submit as a PDF file by e-mail to `lauria at kth dot se` with the subject line `Problem set 4:` ⟨`your name`⟩. Name the PDF file `PS4_`⟨`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 60 points should be enough for grade E, 95 points for grade C, and 130 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) Let $P(x,y)$ be a distribution on $\mathscr{X} \times \mathscr{Y}$ with marginal distributions $P(x)$ and $P(y)$. Show that if $H(Y \mid X) = 0$ then Y is a function of X. That is, show that for all x with $P(x) > 0$ there is only one possible value y such that $P(x,y) > 0$.

2 (10 p) In the multi-party NOF protocol for generalized inner product by Grolmusz that Troy covered in his guest lectures, we viewed the n-bit string inputs to the k players as a $k \times n$-matrix. There was an important subprotocol in Grolmusz's construction that worked provided we had the guarantee for some (fixed but arbitrary) string $r \in \{0,1\}^k$ that r did not appear as a column in this matrix.

Suppose we instead get the guarantee for a particular string r that it appears *at most K times* as a column in the matrix of input string for some small constant K. Does the same approach as in Grolmusz's protocol still work in this setting? If so, do you get any improvements in the upper bound on the communication complexity of generalized inner product by using this idea, and are these improvements substantial or not very significant?

3 (20 p) Suppose G is any DAG with a unique sink and with fan-in bounded by some constant, and consider the lifted pebbling contradiction $\mathit{Lift}_\ell(\mathit{Peb}_G)$ as defined in Massimo's lectures. Prove that if the depth of G — i.e., the length of any longest path from some source vertex to the sink vertex in the graph — is d, then there is a deterministic two-party protocol for the falsified clause search problem $\mathit{Search}(\mathit{Lift}_\ell(\mathit{Peb}_G))$ that has communication cost $\mathrm{O}(d\ell)$.

4 (20 p) Recall that for $x, y \in \{0,1\}^n$ we write $x \leq y$ if $x_i \leq y_i$ for all $i \in [n]$ and say that $f : \{0,1\}^n \mapsto \{0,1\}$ is monotone if $x \leq y$ implies $f(x) \leq f(y)$. Recall also that a monotone Boolean circuit is a circuit C with AND and OR gates but no NOT gates.

4a Prove that any function f computed by a monotone circuit C is monotone. (Note that this proof need not be verbose, but it should be crisp and formally correct.)

4b Prove that for any monotone function f, there is a monotone circuit C that computes f.

4c Is it true that the smallest circuit computing a monotone function is monotone (perhaps up to constant factors in size, say)? Or are there monotone functions for which non-monotone circuits can really help?

For this subproblem, and for this subproblem only, please look at textbooks, search in the research literature, or roam the internet to find an answer. As your solution to this subproblem, state the strongest positive or negative answer to this question you can find together with a solid reference where one can look up a proof (i.e., not a webpage but rather a research paper or possibly textbook). Note that you should still follow the problem set rules in that you are not allowed to collaborate or interact with more than two other persons in order to solve this problem, and these persons should also be participants of this course.

5 (20 p) Prove that the entropy $H(f(X))$ of any function f of X is at most the entropy $H(X)$ of X itself. When does equality hold? Getting the answer right and providing an intuitively convincing explanation will give partial credit, but for full credit a formal proof is also needed.

6 (20 p) Recall the proof of the lower bound on the block sensitivity of the falsified clause search problem for pebbling contradictions Peb_{Π_h} over pyramid graphs Π_h covered in Massimo's lectures. To obtain this lower bound, we considered all paths P from source vertices in Π_h to the sink vertex z and built a "path graph" G with the following properties:

- The vertices $V(G)$ are all source-to-sink paths P.
- There can be an edge (P, Q) only if P and Q start at different source vertices u and v, and if once they intersect at some vertex w they follow exactly the same path from w to z.
- In addition, if (P, Q_1) and (P, Q_2) are edges in G, then it holds that $Q_1 \cap Q_2 \subseteq P$.
- G is undirected, so (P, Q) is an edge if and only if (Q, P) is an edge.

Prove that it holds for such a path graph G that it contains no triangles, i.e., there is no triple $P, Q, R \in V(G)$ such that (P, Q), (P, R), and (Q, R) are all edges in G.

7 (20 p) In this course, we saw essentially the whole proof of the lower bound on the randomized communication complexity of set disjointness $R(\mathsf{DISJ}_n)$ except for the relationship between Hellinger distance $h(P, Q)$ and total variance distance $V(P, Q)$ of two probability distributions P and Q over the same domain. Rectify this omission by proving that $h^2(P, Q) \leq V(P, Q) \leq h(P, Q)\sqrt{2}$ (which is enough to fill in the missing details in the proofs in lectures 10 and 11). *Hint:* Use the facts about Hellinger distance from problem set 3, and do not forget about Cauchy-Schwarz.

8 (30 p) In the last lecture, we presented a randomized reduction of DISJ_n to $\mathsf{KW}_{Match(n,3n)}$. Briefly, given $x, y \subseteq [n]$ Alice and Bob build a pair of graphs G_A and G_B over the same $3n$ vertices such that G_A always has an n-matching but G_B does not. The construction has the property that G_B misses some particular edge e^* that is present in G_A, but if $x \cap y \neq \emptyset$ then in addition at least one more edge in G_A is missing from G_B. We claimed in class that Alice and Bob can now run a deterministic protocol for the problem of finding an edge in $E(G_A) \setminus E(G_B)$, and if $x \cap y \neq \emptyset$ they have at least a 50% chance of detecting this. Namely, this happens if the missing edge that they find is some other edge than e^*.

However, as noted towards the end of the lecture, this claim ignores the following problem: We have no control over what an optimal deterministic protocol for $\mathsf{KW}_{Match(n,3n)}$ looks like. In particular, Alice and Bob build graphs with a very particular structure. Therefore, we could worry that the protocol for some weird reason treats such graphs in a very specific way and always zooms in on exactly the missing edge e^* that we are hoping to avoid.

Investigate what (if anything) can be done to address this concern. For a full score, a rigorous mathematical argument is needed, but getting the answer right and providing an intuitively convincing motivation (regarding whether this is or is not a problem and how it can or cannot be solved) will give partial credit.

9 (30 p) In the very first lecture, we studied an $\mathrm{O}(\log n)$ deterministic two-party protocol for determining the median of $x \cup y$ for $x, y \subseteq [n]$, which very briefly works roughly as follows (see the lecture notes for full details).

Without loss of generality, we assumed for $x = x^{(0)}$ and $y = y^{(0)}$ that $\left|x^{(0)}\right| = \left|y^{(0)}\right| = 2^j$ for some j, and in each round i the protocol halved the sizes of $x^{(i-1)}$ and $y^{(i-1)}$ to get $x^{(i)}$ and $y^{(i)}$ where the median of $x^{(i)} \cup y^{(i)}$ was still the same. We obtained an $\mathrm{O}(\log^2 n)$ protocol by letting Alice and Bob exchange the medians a_i and b_i of their respective sets $x^{(i)}$ and $y^{(i)}$ and then having them throw away half of their elements by using this information. The final optimization step from $\mathrm{O}(\log^2 n)$ to $\mathrm{O}(\log n)$ communication was obtained by observing that Alice and Bob could exchange information about their respective medians a_i and b_i bit by bit starting from the most significant end, since the only important information was whether $a_i < b_i$ or $a_i > b_i$, and that strictly more significant bits for the medians a_{i+k} and b_{i+k} for $k > 0$ in later rounds did not need resending, since a_i and b_i would converge bit by bit to the correct value.

At this point in the lecture we were running short on time, and so the proof of the final optimal protocol was a bit hand-wavy. Later during the course, some students have raised the question how to provide a full, formal proof of correctness of this protocol, and have even worried about whether the claimed invariant about the convergence of the medians a_i and b_i holds.

Investigate what (if anything) can be done to address this concern. For a full score, we want a clear yes/no answer to the question whether there is reason to worry about the invariant or not, and this answer should be backed up by a formal argument. Regardless of whether the answer is yes or no, we also want a rigorous mathematical proof of correctness for an $\mathrm{O}(\log n)$ protocol for the median (which thus might be the protocol sketched above or some modified version of it, depending on whether the protocol needs fixing or not). Just getting the answer right with some kind of intuitively convincing argument (regarding if this is or is not a problem and how it can or cannot be solved) can give partial credit.

10 (40 p) A matching M on an undirected graph $G = (V, E)$ is a subset of edges $(u, v) \in E$ such that every vertex $w \in V$ is mentioned by at most one edge. A *conflict-free* matching M is a matching such that if (u_1, v_1) and (u_2, v_2) are two distinct edges in M, then G does not contain any edge (u_i, v_j) for $i, j \in \{1, 2\}$, $i \neq j$. We say that $G = (V, E)$ is a *k-colour conflict-free matchable graph* if there is a partition of the edges into disjoint colour classes $E = E_1 \cup E_2 \cup \ldots \cup E_k$ such that each E_i is a conflict-free matching in G.

10a Prove that if $G = K_n$ is the complete graph on n vertices with edges between each pair of distinct vertices, then G cannot be k-colour conflict-free matchable for any $k < \binom{n}{2}$.

10b It is a striking combinatorial fact that despite the lower bound on colourable conflict-freeness in Problem 10a, there are almost complete graphs with $\binom{n}{2} - \mathrm{o}(n^2)$ edges that are k-colour conflict-free matchable for $k = n^{1+\mathrm{o}(1)}$ colours.[1]

Recall problem 8 in problem set 2, where Alice, Bob, and Carol get inputs $x, y, z \in \{0, 1\}^n$ and want to determine whether $x = y = z$ or not in the 3-player deterministic number-in-hand message-passing communication model. Use the existence of $n^{1+\mathrm{o}(1)}$-colour conflict-free matchable graphs with $\binom{n}{2} - \mathrm{o}(n^2)$ edges to construct an efficient protocol for this problem. Again, the communication complexity is clearly linear, but we care about getting the best possible multiplicative constant in the upper bound.

Hint: Let the edges in the graph correspond to the bitstrings in Alice's, Bob's and Carol's input.

[1] We write $f(n) = \mathrm{o}(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.