KTH Computer Science and Communication

DD2445 Complexity Theory: Problem Set 3

Submission: Due *Friday December 1, 2017, at 23:59 AoE.* Submit your solutions as a PDF file by e-mail to `jakobn at kth dot se` with the subject line `Problem set 3:` ⟨`your full name`⟩. Name the PDF file `PS3_`⟨`YourFullName`⟩`.pdf` with your name written in CamelCase without blanks and in ASCII without national characters. State your name and e-mail address close to the top of the first page. Solutions should be written in LaTeX or some other math-aware typesetting system with reasonable margins on all sides (at least 2.5 cm). *Please be precise and to the point in your solutions and refrain from vague statements. Write so that a fellow student of yours can read, understand, and verify your solutions.* 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 people are allowed—and indeed, encouraged—but you should always write up your solutions completely on your own, from start to finish, and you should understand all aspects of them fully. It is not allowed to compose draft solutions together and then continue editing individually. You should also clearly acknowledge any collaboration. State close to the top of the first page of your problem set solutions if you have been collaborating with someone and if so with whom. *Note that collaboration is on a per problem set basis, so you should not discuss different problems on the same problem set with different people.*

Reference material: Some of the problems are "classic" and hence 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. Anything said during the lectures or in the lecture notes should be fair game, though, unless you are specifically asked to show something that we claimed without proof in class. All definitions should be as given in class or in Arora-Barak and cannot be substituted by versions from other sources. It is hard to pin down 100% watertight formal rules on what all of this means—when in doubt, ask the main instructor.

About the problems: Some of the problems are meant to be quite challenging and you are not necessarily expected to solve all of them. A total score of 100 points will be enough for grade E, 140 points for grade D, 180 points for grade C, 220 points for grade B, and 260 points for grade A on this problem set. Any corrections or clarifications will be given at `piazza.com/kth.se/fall2017/dd2445/` and any revised versions will be posted on the course webpage `www.csc.kth.se/DD2445/kplx17/`.

1 (10 p) Show that if one-way functions exist, then $\mathsf{P} \neq \mathsf{NP}$.

2 (20 p) Show that $\mathsf{BP} \cdot \mathsf{NP} = \mathsf{AM}[2]$.

3 (20 p) For a function $f : X \times Y \to \{0, 1\}$, let M_f be the matrix of size $|X| \times |Y|$ with rows indexed by $x \in X$ and columns by $y \in Y$ such that $M_f(x, y) = f(x, y)$. Prove that the deterministic 2-party communication complexity of f is bounded from above by $\mathcal{D}^{cc}(f) \leq rank(M_f) + 1$.

Comment: The rank of a matrix can be defined over different fields. You may use without proof the fact that the rank of a matrix over any finite field is bounded from above by the rank over the reals.

4 (30 p) Recall that we defined an *encryption scheme* for plaintexts $x \in \{0,1\}^m$ with encryption keys $k \in \{0,1\}^n$ to be a pair of functions $\bigl(E(k,x), D(k,x)\bigr) = \bigl(E_k(x), D_k(x)\bigr)$ such that for every key k and plaintext x it holds that $D_k(E_k(x)) = x$. An encryption scheme is *perfectly secret* if for every pair of plaintext messages $x, x' \in \{0,1\}^m$ it holds that the distributions $E_{U_n}(x)$ and $E_{U_n}(x')$ are identical (where U_n denotes the uniform distribution over $\{0,1\}^n$).

We claimed in class that no encryption scheme (E,D) with $m > n$ can be perfectly secure. Prove that this is so.
Hint: What happens if all distributions $E_{U_n}(x)$ have the same support?

5 (40 p) Let R_t denote the set of all restrictions of subsets of exactly t out of n variables, where n is supposed to be large and $t \geq n/2$. When proving Håstad's switching lemma in class, we argued that the set $B \subseteq R_t$ of *bad* restrictions for which the conclusion of the lemma does not hold is very small compared to all of R_t, and hence it is very unlikely that a randomly chosen restriction will be bad (which is exactly what the lemma claims).

More formally, we constructed a one-to-one mapping from B to $R_{t+s} \times \{0,1\}^\ell$ for some $\ell = \mathrm{O}(s \log k)$ (where s and k are the parameters in the switching lemma), and claimed this showed that the probability to get a bad restriction is

$$\frac{|B|}{|R_t|} \leq \frac{\bigl|R_{t+s} \times \{0,1\}^\ell\bigr|}{|R_t|} = n^{-\Omega(s)} \ .$$

The purpose of this problem is to fill in the details in these calculations and show that one gets a failure probability for the restriction as small as the one claimed in Håstad's switching lemma *exactly as stated in the textbook Arora-Barak.*

That is, just trusting that the one-to-one map $m : B \to R_{t+s} \times \{0,1\}^\ell$ constructed in class was correct, show that the qoutient $\bigl|R_{t+s} \times \{0,1\}^\ell\bigr| / |R_t|$ is small enough to give the probability bound in the switching lemma as stated in the textbook.
Hint: Show that for $t > n/2$ it holds that

$$\binom{n}{t+s} \leq \binom{n}{t} \left(\frac{e(n-t)}{n}\right)^s$$

by first proving

$$\binom{n}{t+s} = \binom{n}{t}\binom{n-t}{s} \Big/ \binom{t+s}{t}$$

(and try to find a nice combinatorial proof for this latter equality). You can use the well-known inequalities

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

without proof.

6 (70 p) The purpose of this problem is to investigate some of the conditions in Håstad's switching lemma, in particular, the requirement of bounded fan-in (i.e., that the restrictions operate on k-CNF and k-DNF formulas).

6a (25 p) Let $f : \{0,1\}^n \to \{0,1\}$ be some Boolean function. Prove that if all (minimal) maxterms of f have size at most s, then f can be represented as an s-CNF formula.

Does the other direction hold as well? That is, is it true that if f can be represented as an s-CNF formula then all (minimal) maxterms of f have size at most s?

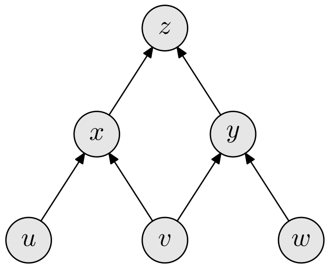
6b (25 p) Prove that any CNF formula that computes parity of n bits must have size exponential in n. For full credit, prove an exact, tight bound. (And for concreteness, define the *size* of a CNF formula as the number of literals in it, counted with repetitions).

6c (20 p) Argue that in view of Problem 6b, we actually do not need the added requirement of bounded fan-in in the final step of the proof of PARITY $\notin \mathsf{AC}^0$, i.e., after $(d-2)$ rounds of restrictions have been applied on C' so that the circuit has collapsed to a CNF formula. In our proof in class, we crucially used in this step that the formula obtained was a k'-CNF formula for some constant k'. (Let us note in passing that there is a lower bound on DNF formulas analogous to that in Problem 6b in case the circuit collapses to a DNF formula, but there is no need to prove this or even consider the DNF case to get a full score.)

This raises the question whether we could in fact drop the restriction on fan-in in the bottom layer completely at all $(d-2)$ stages of the proof if we just did a little bit of extra work. Explain how to modify the proof of PARITY $\notin \mathsf{AC}^0$ to work also if there is no bound on the bottom-level fan-in of C' (if this can be done), or point out where in the proof we run into trouble (if it cannot be done).

7 (50 p) Let multiprover interactive protocols be defined as the interactive protocols in Section 8.1 in Arora-Barak, except that there are several provers and that the verifier's messages in each round depends on previous messages from all provers (and on the verifier's private randomness). The messages sent by each prover only depends on the communication with the verifier, however, just as before. Let $\mathsf{MIP}[N]$ denote the set of languages that can be decided by N-multiprover interactive protocols in a polynomial number of rounds (in analogy with $\mathsf{IP} = \mathsf{MIP}[1]$ in Definition 8.6 in Arora-Barak).

Prove that, as claimed in class, only two provers are needed to realize the full power of multiprover interactive protocols. That is, prove that $\mathsf{MIP}[2] = \mathsf{MIP}[\text{poly}]$, where $\mathsf{MIP}[\text{poly}]$-protocols have a number of provers scaling polynomially with the size of the input.



(a) Pyramid graph Π_2 of height 2.

$$\begin{aligned} & u \\ \wedge\ & v \\ \wedge\ & w \\ \wedge\ & (\overline{u} \vee \overline{v} \vee x) \\ \wedge\ & (\overline{v} \vee \overline{w} \vee y) \\ \wedge\ & (\overline{x} \vee \overline{y} \vee z) \\ \wedge\ & \overline{z} \end{aligned}$$

(b) Pebbling contradiction Peb_{Π_2}.

Figure 1: Example pebbling contradiction for the pyramid of height 2.

8 (50 p) The *falsified clause search problem* is the following communication problem. The starting point is some fixed unsatisfiable CNF formula F and some fixed partition $X \dot{\cup} Y = \mathit{Vars}(F)$ of the variables of F between Alice and Bob. Given as inputs truth value assignments $\alpha_X : X \rightarrow \{0,1\}$ and $\alpha_Y : Y \rightarrow \{0,1\}$, Alice and Bob should communicate to find a clause $C \in F$ that is falsified by the assignment $\alpha = \alpha_X \cup \alpha_Y$. (Such a clause always exists since F is unsatisfiable.)

The *pyramid graph* Π_h of height h is a DAG with $h+1$ layers, where there is one vertex in the highest layer (the sink z), two vertices in the next layer et cetera, down to $h+1$ vertices in the lowest layer 0. The ith vertex in layer L has incoming edges from the ith and $(i+1)$st vertices in layer $L-1$. See Figure 1a for an illustration of the pyramid graph of height 2.

The purpose of this problem is to investigate the hardness of the falsified clause search problems for certain CNF formulas defined in terms of pyramids.

8a (30 p) The so-called *pebbling formula* over Π_h, denoted Peb_{Π_h}, is the conjunction of the following clauses:

- for all vertices s in the bottom layer, a unit clause s (i.e., a clause of size 1),
- For all vertices w in layers $L \geq 1$ with predecessors u, v, the clause $\overline{u} \vee \overline{v} \vee w$,
- for the sink z, the unit clause $\overline{z}$.

Figure 1b shows the formula corresponding to the pyramid in Figure 1a.

Give the best upper and lower bounds you can for the deterministic communication complexity of the falsified clause search problem for Peb_{Π_h}. Your bounds should hold for any (arbitrary but fixed) partition $X \dot{\cup} Y = \mathit{Vars}\big(Peb_{\Pi_h}\big)$ of the variables. Express your bounds in terms of the number of vertices $n = (h+1)(h+2)/2$ in the graph Π_h. For full credit the bounds should be asymptotically tight.

$$
\begin{array}{ll}
\ (u_1 \vee u_2) & \wedge\ (v_1 \vee \overline{v}_2 \vee \overline{w}_1 \vee w_2 \vee y_1 \vee y_2) \\
\wedge\ (\overline{u}_1 \vee \overline{u}_2) & \wedge\ (v_1 \vee \overline{v}_2 \vee \overline{w}_1 \vee w_2 \vee \overline{y}_1 \vee \overline{y}_2) \\
\wedge\ (v_1 \vee v_2) & \wedge\ (\overline{v}_1 \vee v_2 \vee w_1 \vee \overline{w}_2 \vee y_1 \vee y_2) \\
\wedge\ (\overline{v}_1 \vee \overline{v}_2) & \wedge\ (\overline{v}_1 \vee v_2 \vee w_1 \vee \overline{w}_2 \vee \overline{y}_1 \vee \overline{y}_2) \\
\wedge\ (w_1 \vee w_2) & \wedge\ (\overline{v}_1 \vee v_2 \vee \overline{w}_1 \vee w_2 \vee y_1 \vee y_2) \\
\wedge\ (\overline{w}_1 \vee \overline{w}_2) & \wedge\ (\overline{v}_1 \vee v_2 \vee \overline{w}_1 \vee w_2 \vee \overline{y}_1 \vee \overline{y}_2) \\
\wedge\ (u_1 \vee \overline{u}_2 \vee v_1 \vee \overline{v}_2 \vee x_1 \vee x_2) & \wedge\ (x_1 \vee \overline{x}_2 \vee y_1 \vee \overline{y}_2 \vee z_1 \vee z_2) \\
\wedge\ (u_1 \vee \overline{u}_2 \vee v_1 \vee \overline{v}_2 \vee \overline{x}_1 \vee \overline{x}_2) & \wedge\ (x_1 \vee \overline{x}_2 \vee y_1 \vee \overline{y}_2 \vee \overline{z}_1 \vee \overline{z}_2) \\
\wedge\ (u_1 \vee \overline{u}_2 \vee \overline{v}_1 \vee v_2 \vee x_1 \vee x_2) & \wedge\ (x_1 \vee \overline{x}_2 \vee \overline{y}_1 \vee y_2 \vee z_1 \vee z_2) \\
\wedge\ (u_1 \vee \overline{u}_2 \vee \overline{v}_1 \vee v_2 \vee \overline{x}_1 \vee \overline{x}_2) & \wedge\ (x_1 \vee \overline{x}_2 \vee \overline{y}_1 \vee y_2 \vee \overline{z}_1 \vee \overline{z}_2) \\
\wedge\ (\overline{u}_1 \vee u_2 \vee v_1 \vee \overline{v}_2 \vee x_1 \vee x_2) & \wedge\ (\overline{x}_1 \vee x_2 \vee y_1 \vee \overline{y}_2 \vee z_1 \vee z_2) \\
\wedge\ (\overline{u}_1 \vee u_2 \vee v_1 \vee \overline{v}_2 \vee \overline{x}_1 \vee \overline{x}_2) & \wedge\ (\overline{x}_1 \vee x_2 \vee y_1 \vee \overline{y}_2 \vee \overline{z}_1 \vee \overline{z}_2) \\
\wedge\ (\overline{u}_1 \vee u_2 \vee \overline{v}_1 \vee v_2 \vee x_1 \vee x_2) & \wedge\ (\overline{x}_1 \vee x_2 \vee \overline{y}_1 \vee y_2 \vee z_1 \vee z_2) \\
\wedge\ (\overline{u}_1 \vee u_2 \vee \overline{v}_1 \vee v_2 \vee \overline{x}_1 \vee \overline{x}_2) & \wedge\ (\overline{x}_1 \vee x_2 \vee \overline{y}_1 \vee y_2 \vee \overline{z}_1 \vee \overline{z}_2) \\
\wedge\ (v_1 \vee \overline{v}_2 \vee w_1 \vee \overline{w}_2 \vee y_1 \vee y_2) & \wedge\ (z_1 \vee \overline{z}_2) \\
\wedge\ (v_1 \vee \overline{v}_2 \vee w_1 \vee \overline{w}_2 \vee \overline{y}_1 \vee \overline{y}_2) & \wedge\ (\overline{z}_1 \vee z_2)
\end{array}
$$

Figure 2: XORified pebbling contradiction $Peb_{\Pi_2}[\oplus]$.

8b (20 p) In the *XORified pebbling formula* $Peb_{\Pi_h}[\oplus]$ over Π_h we instead think of each vertex v as the exclusive or of two variables $v_1 \oplus v_2$ and have the following clauses:

- for all vertices s in the bottom layer, the CNF encoding of $s_1 \oplus s_2$,
- For all w in layers $L \geq 1$ with predecessors u, v, the CNF encoding of $\neg(u_1 \oplus u_2) \vee \neg(v_1 \oplus v_2) \vee (w_1 \oplus w_2)$,
- for the sink z, the CNF encoding of $\neg(z_1 \oplus z_2)$.

Figure 2 shows the XORified pebbling formula for the pyramid in Figure 1a.

Give the best upper bound you can for the deterministic communication complexity of the falsified clause search problem for $Peb_{\Pi_h}[\oplus]$, where Alice gets all variables $X = \{u_1, v_1, w_1, \ldots\}$ and Bob gets all variables $Y = \{u_2, v_2, w_2, \ldots\}$ (again expressed in terms of the number of vertices $n = (h+1)(h+2)/2$ in Π_h).

8c (∞ p) Prove a $\omega(\log n)$ deterministic communication complexity lower bound for the falsified clause search problem for $Peb_{\Pi_h}[\oplus]$ with variables partitioned as in Problem 8b (ideally a bound on the form $\Omega(n^\delta)$ for some $\delta > 0$), or establish that no such lower bound exist.

Remark: This is an open research problem, and so you are not necessarily expected to solve it...

9 (60 p) Consider a set $A \subseteq \mathcal{A}^n$ where we interpret $a \in A$ as a vector of length n with an element from $\mathcal{A}$ in each coordinate a_i. For each $i \in [n]$, define the bipartite graph $G(A, i) = (U_i \dot\cup V_i, E)$ to have left vertex set $U_i = \{a_i \mid a \in A\}$ and right vertex set $V_i = \{a_{\neq i} \mid a \in A\}$, where $a_{\neq i}$ denotes the vector of length $n-1$ obtained by omitting the coordinate a_i from a, and to have an edge between $a_i \in U_i$ and $a_{\neq i} \in V_i$ for each $a \in A$. Let us define the quantities

$$d_{\min}(A, i) = \min_{v \in V_i}\{|\mathsf{Ext}(v)|\}$$

and

$$d_{\mathrm{avg}}(A, i) = \frac{|A|}{|A_{\neq i}|} \ ,$$

where $\mathsf{Ext}(v) = \{a \in A \mid a_{\neq i} = v\}$. Prove that for $n \geq 2$ it holds that if $d_{\mathrm{avg}}(A, i) \geq K \cdot |\mathcal{A}|$ for all $i \in [n]$, then there is a subset $A' \subseteq A$ with $|A'| \geq \frac{|A|}{2}$ such that $d_{\min}(A', i) \geq \frac{K}{2n} \cdot |\mathcal{A}|$ for all $i \in [n]$.

Comment: This is Lemma 13.4 from Lecture 13.

10 (60 p) The goal of this exercise is to give a complete proof that $\mathsf{PSPACE} \subseteq \mathsf{IP}$, strengthening the result $\mathsf{coNP} \subseteq \mathsf{IP}$ that was proven in class.

Given a quantified Boolean formula (QBF) $\psi = \forall x_1 \exists x_2 \forall x_3 \cdots \exists x_n\, \phi(x_1, \ldots, x_n)$, we can use arithmetization as in our proof of $\mathsf{coNP} \subseteq \mathsf{IP}$ to construct a polynomial P_ϕ such that ψ is true if and only if $\prod_{b_1 \in \{0,1\}} \sum_{b_2 \in \{0,1\}} \prod_{b_3 \in \{0,1\}} \cdots \sum_{b_n \in \{0,1\}} P_\phi(b_1, \ldots, b_n) \neq 0$. However, the SumCheck protocol we used to decide the $\#\mathrm{SAT}_D$ problem for CNF formulas no longer works, since each multiplication corresponding to a $\forall$-quantifier can double the degree of the polynomial.

10a (20 p) Suppose that ψ is a QBF formula (not necessarily in *prenex normal form* as described in Definition 4.10 and discussed further below in Arora-Barak) satisfying the following property: if $x_1, \ldots, x_n$ are the variables of ψ sorted in order of first appearance, then for every variable x_i there is at most a single universal quantifier involving x_j for any $j > i$ appearing before the last occurrence of x_i in ψ. Show that in this case, when we run the SumCheck protocol with the modification that we check $s(0) \cdot s(1) = K$ for product operations (i.e., $\forall$-quantifiers), the prover only needs to send polynomials of degree $\mathrm{O}(n)$ since the degree blow-up is at most a constant factor 2.

10b (20 p) Assuming that any QBF formula ψ can be rewritten to satisfy the property in Problem 10a, use this to show that $\mathrm{TQBF} \in \mathsf{IP}$ and hence $\mathsf{PSPACE} \subseteq \mathsf{IP}$.

10c (20 p) Show that any QBF formula ψ of size m can be transformed into a logically equivalent formula ψ' of size $\mathrm{O}(m^2)$ that satisfies the property in Problem 10a.

Hint: Introduce a new variable y_i for any occurrence of x_i that we need to get rid of and encode that x_i and y_i take the same truth value.