Problem 1 (5 points) Let R be a relation with attributes ABCD. Reduce the number of joins in the algebra query
\[ \pi_{AB}[\pi_{BCD}(R) \bowtie \pi_{ACD}(R)] \bowtie \pi_{AD}(R) \]
knowing that the query is only applied to databases satisfying the dependencies:
\[ B \rightarrow D, D \rightarrow C, A \rightarrow B. \]

Start by writing a CQ equivalent to the above algebra query, then chase and minimize its body, and finally convert the result back to an algebra query.

Problem 2 (4 points) Show that there is no algorithm that, given a relational algebra query, returns an equivalent query with a minimum number of joins.

Problem 3 (10 points) A database holds data about courses in an undergraduate CSE program, prerequisites, and records for each student. The tables are the following:

<table>
<thead>
<tr>
<th>Course</th>
<th>C-id</th>
<th>Record</th>
<th>S-id</th>
<th>C-id</th>
<th>Prereq</th>
<th>C-id</th>
<th>P-id</th>
</tr>
</thead>
</table>

In relation Prereq, P-id provides a prerequisite for C-id (assume this is acyclic). Relation Record provides, for each student S-id, the courses C-id he/she has already taken. Assume that Record[C-id] ⊆ Course[C-id], Prereq[C-id] ⊆ Course[C-id] and Prereq[P-id] ⊆ Course[C-id]. Consider the following query:

“For each student, find the set of courses whose prerequisites he/she has already taken”

Express this query in the following languages: (i) CALC (ii) relational algebra (iii) nr-Datalog¬ (iv) inflationary Datalog¬, and (v) semi-positive Datalog¬, assuming also given a successor relation succ on course ids, as well as constants min and max providing the minimum, resp. maximum course id.

Problem 4 (5 points) Let G be a binary relation representing the edges of a directed graph. (i) (3 points) Is there a while⁺ program defining the property
that the diameter of $G$ is a perfect square? If yes, explain the idea *without* attempting to explicitly construct such a program. (ii) (2 points) Consider the set $G$ of directed graphs $G$ in which all nodes have in-degree and out-degree 1. Does $\text{while}^+$ express all \text{PTIME} queries on $G$?

**Problem 5** (2 points) Let $G$ be a binary relation representing the edges of a directed graph. Is there a CALC sentence stating that $G$ consists of a single cycle?

**Problem 6** (3 points) We know that Datalog queries are monotonic and can be evaluated in PTIME (with respect to the database). Does Datalog express all the monotonic PTIME queries?

**Problem 7** (6 points) (i) (3 points) Prove that it is undecidable whether a given $\text{while}^+$ program is equivalent to some CALC query. (ii) (3 points) Consider a database consisting of a unary relation $R$ (i.e., $R$ has just one attribute). Prove that every property of $R$ definable by a $\text{while}^+$ program with no constants is also definable in CALC. **Hint:** there is a 3-line proof.