Problem 1 (10 marks) Consider a sound GAV setting with source relations \( R_1(A, B) \) and \( R_2(B, C) \) and global schema relations \( G_1(A, C) \) and \( G_2(B, C) \) with the mapping defined by

\[
G_1 \supseteq \pi_{AC}(R_1 \bowtie B R_2) \\
G_2 \supseteq \pi_{BC}(\sigma_{A=1}(R_1 \bowtie B R_2))
\]

Consider a query \( Q = G_1 \bowtie C G_2 \). Show how to compute certain answers to \( Q \) using the sources. Why does this solution work? Write your solution as an SQL query (in terms of the source relations).

Problem 2 (20 marks) We mentioned in class that

\( \text{(*)} \) there is no algorithm that checks, for a relational algebra expression \( e \), whether \( e(D) = \emptyset \) for every possible database \( D \).

To show that query answering in LAV or GAV data exchange is undecidable (impossible to compute algorithmically) for relational algebra queries, we need a slightly different assumption: there is no algorithm that checks whether the result of a relational algebra expression is constant, i.e. independent of the input database.

Your goal is to prove this statement under the assumption (*)

Note that correctness of a proof is often inversely proportional to its length – verbosity rarely translates into correctness. If you go beyond 10-15 lines, it probably means something is seriously wrong!

Problem 3 (30 marks) This question is about optimization of conjunctive queries. Consider two SQL queries below over relations \( R(A, B) \) and \( S(B, C) \).

Query \( Q_1 \)

```
SELECT R1.A, S1.C
FROM R R1, R R2, R R3, S S1 S S2
```

Query \( Q_2 \)

```
SELECT R2.A, S2.C
FROM R R1, R R2, S S1 S S2
```

Answer the following questions. Each one is worth 10 marks.

1. Write both \( Q_1 \) and \( Q_2 \) as rule-based queries.

2. Is \( Q_1 \) contained in \( Q_2 \)? Is \( Q_2 \) contained in \( Q_1 \)? Explain your answer.

3. Find a query equivalent to \( Q_1 \) that has the minimum number of joins. Express it both as an SQL query and as a relational algebra query.
Problem 4 (40 marks) This question is about LAV (local-as-view) data integration. We have a global schema with two relations $G_1(A, B)$ and $G_2(B, C)$ and two sources $S_1$ and $S_2$ such that the LAV mapping is provided by the SQL queries below:

```sql
FROM G1, G2
WHERE G1.B=G2.B
```

```sql
SELECT G2.C
FROM G1, G2
WHERE G1.B=G2.B
```

That is, the content of the first source is the result of applying the first query to a global-schema database, and likewise for the second query.

In addition we have a query $Q$ over the global schema given by:

```sql
SELECT G11.A
FROM G1 G11, G1 G12, G1 G13, G1 G14, G2 G21, G2 G22, G2 G23
```

The goal is to see how $Q$ can be answered over the sources. To do so, you must answer the following questions. The first and the third are worth 10 marks, the second is worth 20 marks.

1. Express the views defining $S_1$ and $S_2$, as well as the query $Q$, as rule-based queries and as tableaux.
2. Find a rewriting of $Q$ using $S_1$ and $S_2$. Explain how you achieve it; in this step it suffices to provide a rewriting as a tableau or a rule-based query.
3. Express the rewriting from the previous item in both relational algebra and SQL.