1. The movie database consists of the following two relations

\[ \text{movie: title, director, actor} \]
\[ \text{schedule: theater, title} \]

The first relation provides titles, directors, and actors of various movies. Assume a movie is uniquely identified by its title. Do not assume that a movie has a unique director (so a movie by Hitchcock is one for which one of the directors is Hitchcock). The second relation provides the titles of currently playing movies and the theaters where they are being shown. Express the following queries in (i) relational calculus and (ii) relational algebra.

(a) List the theaters showing some movie by Hitchcock.

(i) \( \{ t : th \mid \exists x \exists a ( \text{schedule}(t, x) \land \text{movie}(x, \text{Hitchcock}, a)) \} \)

(ii) \( \pi_{th}(\text{schedule}) \land \pi_{title}(\sigma_{\text{director}=\text{Hitchcock}(\text{movie}))} \)

(b) List the theaters showing only movies by Hitchcock.

(i) \( \{ t : th \mid \exists x \exists a (\text{sch}(t, x)) \land \forall x [\text{sch}(t, x) \rightarrow \exists a (\text{movie}(x, \text{Hitchcock}, a))] \} \)

With active domain semantics, the conjunct \( \exists x (\text{sch}(t, x)) \) is needed to make sure \( t \) is a theater.

(ii) \( \pi_{th}(\text{sch}) - \pi_{th}[\text{sch} \land (\pi_{title}(\text{movie}) - \pi_{title}(\sigma_{\text{dir}=\text{Hitchcock}(\text{movie}))])] \)

2. The division binary operator \( \div \) on relations is defined as follows. Given relations \( P \) and \( Q \) for which \( \text{att}(Q) \subset \text{att}(P) \), \( P \div Q \) is a relation with attributes \( \text{att}(P) - \text{att}(Q) \) containing the tuples \( t \) for which \( \{ t \} \bowtie Q \subseteq P \). For example, if \( \text{att}(P) = \{ A, B \} \) and \( \text{att}(Q) = \{ B \} \), \( P \div Q \) is the relation with attribute \( \{ A \} \) containing the tuples \( \langle a \rangle \) for which \( \langle a, b \rangle \in P \) for every tuple \( \langle b \rangle \in Q \). Intuitively, \( \div \) is a direct implementation of universal quantification.
(i) Use ÷ (and standard algebra operators) to express the query “List the theaters showing every movie by Hitchcock”. 

\[ \text{schedule} ÷ \pi_{\text{title}}(\sigma_{\text{dir}=\text{Hitch}}(\text{movie})) \]

(ii) Show how \( P ÷ Q \) can be expressed using the standard relational algebra operators (you can assume, for simplicity, that \( \text{att}(P) = \{A, B\} \) and \( \text{att}(Q) = \{B\} \)).

\[ \pi_A(P) - \pi_A(\pi_A(P) \bowtie Q - P) \]

3. Consider the following query on the above schedule relation:

Find the theaters showing more than one title

(i) Express this query in relational calculus and relational algebra.

\[ \{ t : \text{th} \mid \exists x_1 \exists x_2 (\text{sch}(t, x_1) \land \text{sch}(t, x_2) \land x_1 \neq x_2) \} \]

\[ \pi_{\text{th}}(\sigma_{\text{title} \neq \text{title}_1}(\text{schedule} \bowtie \delta_{\text{title} \rightarrow \text{title}_1}(\text{schedule}))) \]

\[ (\sigma_{A \neq B}(e)) \text{ stands for } e - \sigma_{A=B}(e). \]

(ii) (*) Prove that every relational algebra expression defining the above query must use the attribute renaming operator \( \delta \).

Suppose \( e \) expresses the query and uses no \( \delta \). In the proof below we also assume that \( e \) uses no constant relations (the proof with constant relations is slightly more complicated but follows the same lines). However, \( e \) may use selection with constants.

Let \( I \) be the instance of \( \text{sch} \):

\[
\begin{array}{c|cc}
\text{sch} & \text{th} & \text{title} \\
\hline
a & 1 & \text{} \\
a & 2 & \text{} \\
b & 3 & \text{}
\end{array}
\]

where \( e \) does not use selection with any of the constants in \( C = \{a, b, 1, 2, 3\} \). Consider any relational algebra expression \( f \) over \( \text{sch} \)
using no renaming and no constants in $C$ ($e$ is one such expression). Denote by $\text{att}(f)$ the set of attributes of the result of $f$. Obviously, $\text{att}(f) \subseteq \{\text{th}, \text{title}\}$ since no renaming is used. We denote by $\bowtie I$ the relation $\pi_{\text{th}}(I) \bowtie \pi_{\text{title}}(I)$ and $\bar{I} = \bowtie(I) - I$. We show by structural induction the following:

- if $f$ has one attribute $A \in \{\text{th}, \text{title}\}$, then $f(I) = \emptyset$ or $f(I) = \pi_A(I)$.
- if $\text{att}(f) = \{\text{th}, \text{title}\}$ then $f(I) \in \{\emptyset, I, \bar{I}, \bowtie(I)\}$

Note that this shows that $e$ cannot express the desired query, since $\text{att}(e) = \{\text{th}\}$ so by the above $e(I)$ must equal $\emptyset$ or $\pi_{\text{th}}(I)$, neither of which is the desired answer $\{a\}$.

**Basis** $f = \text{sch}$. Then $\text{att}(f) = \{\text{th}, \text{title}\}$ and $f(I) = I$.

**Induction step** We consider the following cases, where $f_1$ and $f_2$ are algebra expressions over $\text{sch}$ assumed, by the induction hypothesis, to satisfy the above statement:

- $f = \sigma_{A=c}(f_1)$. Then $f(I) = \emptyset$, since $c \notin C$, so $f_1(I)$ does not contain $c$.
- $f = \sigma_{A=B}(f_1)$. Then $\text{att}(f_1) = \{\text{th}, \text{title}\}$, so $A = \text{th}$ and $B = \text{title}$ (or conversely). By the induction hypothesis, $f_1(I) \in \{\emptyset, I, \bar{I}, \bowtie(I)\}$, so $\sigma_{A=B}(f_1(I)) = \emptyset$.
- $f = \pi_A(f_1)$. We can assume wlog$^1$ that $\text{att}(f_1) = \{\text{th}, \text{title}\}$. By the induction hypothesis, $f_1(I) \in \{\emptyset, I, \bar{I}, \bowtie(I)\}$. It follows that $f(I)$ is either empty or equals $\pi_A(I)$.
- $f$ is $f_1 \bowtie f_2$, $f_1 \cup f_2$, or $f_1 - f_2$. By the induction hypothesis, $f_j(I) \in \{\emptyset, \pi_{\text{th}}(I), \pi_{\text{title}}(I), I, \bar{I}, \bowtie(I)\}$. By considering all possibilities (subject to typing constraints), we see that $f(I)$ satisfies the statement.

This completes the induction and proves (ii).

4. (Automorphisms)

(i) (5 points) Show that CALC queries are invariant under automorphisms. In other words, if $\varphi(\bar{x})$ is a CALC query without constants

---

$^1$Without loss of generality.
Let $\varphi(\bar{x})$ be a CALC formula with free variables $\bar{x}$. We prove the following:

1. For each instance $I$ and automorphism $f$ of $I$, $f(\varphi(I)) \subseteq \varphi(I)$.

Once (1) is proven, the converse inclusion $\varphi(I) \subseteq f(\varphi(I))$ is shown as follows. Since $f$ is an automorphism of $I$, so is $f^{-1}$. By (1) we have that $f^{-1}(\varphi(I)) \subseteq \varphi(I)$. But then $f(f^{-1}(\varphi(I))) \subseteq f(\varphi(I))$, so $\varphi(I) \subseteq f(\varphi(I))$.

The proof of (1) is by structural induction on the formula. Suppose $\varphi(\bar{x})$ is an atom. If $\varphi(\bar{x}) = R(\bar{x})$ then $\varphi(I) = I(R)$, so $f(\varphi(I)) = f(I(R)) = I(R) = \varphi(I)$. If $\varphi$ is $x = y$ then $\varphi(I) = \{\langle a, a \rangle \mid a \in \text{dom}(I)\}$ so $f(\varphi(I)) = \{\langle f(a), f(a) \rangle \mid a \in \text{dom}(I)\} = \varphi(I)$ (since $f(\text{dom}(I)) = \text{dom}(I)$).

For the induction step, we have three cases:

1. $\varphi = \varphi(\bar{x}, \bar{y}, \bar{z}) = \alpha(\bar{x}, \bar{y}) \land \beta(\bar{y}, \bar{z})$. Let $\langle f(\bar{u}), f(\bar{v}), f(\bar{w}) \rangle \in f(\varphi(I)),$ with $\langle \bar{u}, \bar{v}, \bar{w} \rangle \in \varphi(I)$. Then $\langle \bar{u}, \bar{v}, \bar{w} \rangle \in \alpha(I)$ and $\langle \bar{v}, \bar{v}, \bar{w} \rangle \in \beta(I)$. Since $f(\alpha(I)) = \alpha(I)$, $\langle f(\bar{u}), f(\bar{v}) \rangle \in \alpha(I)$ and similarly $\langle f(\bar{v}), f(\bar{w}) \rangle \in \beta(I)$. It follows that $\langle f(\bar{u}), f(\bar{v}), f(\bar{w}) \rangle \in \varphi(I)$. We conclude that $f(\varphi(I)) \subseteq \varphi(I)$.

2. $\varphi = \exists x \alpha(x, \bar{y})$. Let $\langle \bar{u} \rangle \in \varphi(I)$. Then there exists $v \in \text{dom}(I)$ such that $\langle v, \bar{u} \rangle \in \alpha(I)$. Since $f(\alpha(I)) = \alpha(I)$, it follows that $\langle f(v), f(\bar{u}) \rangle \in \alpha(I)$, and $f(\bar{u}) \in \varphi(I)$. Thus, $f(\varphi(I)) \subseteq \varphi(I)$.

3. $\varphi = \neg \alpha(\bar{x})$. Let $\bar{u} \in \varphi(I)$. Then $\bar{u} \in \text{dom}(I)$ and $\bar{u} \notin \alpha(I)$. But then $f(\bar{u}) \in \text{dom}(I)$ and $f(\bar{u}) \notin \alpha(I)$. Indeed, suppose $f(\bar{u}) \in \alpha(I)$. Since $f^{-1}$ is an automorphism of $I$, $f^{-1}(\alpha(I)) \subseteq \alpha(I)$ by the induction hypothesis, so $f^{-1}(f(\bar{u})) \in \alpha(I)$ and $\bar{u} \in \alpha(I)$. However, this is false. We conclude that $f(\bar{u}) \notin \alpha(I)$, so $f(\bar{u}) \in \varphi(I)$ and $f(\varphi(I)) \subseteq \varphi(I)$.

This completes the proof of (1) and of (i).

(ii) (2 points) Let $\sigma$ consist of a binary relation $R$. Using (i), show that there is no CALC query without constants which on input
produces as answer

\[ R \]
\[
\begin{array}{cc}
  a & b \\
  b & c \\
  c & d \\
  d & a \\
\end{array}
\]

Suppose there is a CALC query \( \varphi \) with no constants, with the above property. Let \( I \) be the input in the figure. By (i), \( f(\varphi(I)) = \varphi(I) \) for every automorphism \( f \) of \( I \). Consider the 1-1 mapping \( f \) on \( \text{dom}(I) \) defined by the table

\[ f \]
\[
\begin{array}{c}
  a \\
  b \\
  c \\
  d \\
\end{array} \quad \begin{array}{c}
  \rightarrow \\
  \rightarrow \\
  \rightarrow \\
  \rightarrow \\
\end{array} \quad 
\begin{array}{c}
  b \\
  c \\
  d \\
  a \\
\end{array}
\]

Clearly \( f(I) = I \). However, \( f(\varphi(I)) \neq \varphi(I) \), a contradiction.