# Retake Examination in the Course Automated Theorem Proving

Prof. Dr. Jasmin Blanchette, Lydia Kondylidou,
Yiming Xu, PhD, and Tanguy Bozec

Winter Term 2024/25
3 April 2025

**For convenience, a handout is provided with the definitions of the main calculi and concepts covered in the course.**

**Last name (in CAPITAL LETTERS):**

**First name (in CAPITAL LETTERS):**

**Matriculation number:**

**Program of study:**

☐ **Please check with an X *only* if the exam should be voided and not graded.**

**Bitte *nur* ankreuzen, wenn die Klausur entwertet und nicht korrigiert werden soll.**

**If your program of study does not include a voiding rule, an X mark will lead to a failing grade.**

Hereby I confirm the correctness of the above information: ____________________

Signature

With your signature, you confirm that you are in sufficiently good health at the beginning of the examination and that you accept this examination bindingly.

Please leave the following table blank:

| Question | 1 | 2 | 3 | 4 | 5 | 6 | $\sum$ |
|---|---|---|---|---|---|---|---|
| Points | 18 | 16 | 14 | 18 | 16 | 18 | 100 |
| Score | | | | | | | |

## Instructions

You have **120 minutes** at your disposal. Written or electronic aids (except normal watches) are not permitted. Carrying electronic devices, even switched off, will be considered cheating.

Write your full name and matriculation number clearly legible on the cover sheet, as well as your name in the header on each sheet. Hand in all sheets. Leave them stapled together. Use only **pens** and **neither** the color **red nor green**.

Check that you have received all the sheets and the handout. The questions can be found on **pages 3–11**. You may use the back of the sheets for secondary calculations. If you use the back of a sheet to answer, clearly mark what belongs to which question and indicate in the corresponding question where all parts of your answer can be found. Cross out everything that should not be graded.

There are 6 questions for a total of 100 points.

**Answer.** This version of the exam contains model answers in blocks like this one.

**Grading.** And blocks like this one specify the grading scheme.

**Question 1** (18 points)**:** The first-order formula $F = \exists x\,(D(x) \rightarrow \forall y\, D(y))$ over $\Sigma = (\{b/0\}, \{D/1\})$ is called the *drinker formula*. It is valid. Informally: "There exists a person $x$ such that if $x$ drinks, then everybody drinks."

(a) Use the (unordered) resolution calculus to prove $F$. You will first need to transform $F$ first to obtain a problem in clausal normal form. Spell out all your steps.

**Answer.** (a) We start with the CNF transformation:

$$
\begin{array}{rl}
 & \neg\exists x\,(D(x) \rightarrow \forall y\, D(y)) \\
\Rightarrow_P & \forall x\,\neg(D(x) \rightarrow \forall y\, D(y)) \\
\Rightarrow_P & \forall x\,\neg\forall x_1\,(D(x) \rightarrow D(x_1)) \\
\Rightarrow_P & \forall x\,\exists x_1\,\neg(D(x) \rightarrow D(x_1)) \\
\Rightarrow_S & \forall x\,\neg(D(x) \rightarrow D(f_1(x))) \\
\Rightarrow_{CNF} & \forall x\,\neg(\neg D(x) \vee D(f_1(x))) \\
\Rightarrow_{CNF} & \forall x\,(\neg\neg D(x) \wedge \neg D(f_1(x))) \\
\Rightarrow_{CNF} & \forall x\,(D(x) \wedge \neg D(f_1(x)))
\end{array}
$$

Thus $N = \{D(x),\ \neg D(f_1(x))\}$.

A single "Resolution" inference from these two clauses (in which the variables are renamed apart) generates the empty clause.

**Grading.** 12 points:

- 10 points for clausification
  - 4 points for the result of clausification
  - 1 points for each of the six intermediate steps
- 2 points for the resolution inference

(b) Use the tableau calculus with ground instantiation to prove $F$. Use exactly the expansion rules seen in the lecture. Document all your steps.

**Answer.**

| | | |
|---|---|---|
| 1. | $\neg\exists x\,(D(x) \to \forall y\, D(y))$ | $F$'s negation |
| 2. | $\neg(D(b) \to \forall y\, D(y))$ | $1(b)[\gamma]$ |
| 3. | $D(b)$ | $2_1[\alpha]$ |
| 4. | $\neg\forall y\, D(y)$ | $2_2[\alpha]$ |
| 5. | $\neg D(c_1)$ | $4[\delta]$ |
| 6. | $\neg(D(c_1) \to \forall y\, D(y))$ | $1(c_1)[\gamma]$ |
| 7. | $D(c_1)$ | $6_1[\alpha]$ |
| 8. | $\neg\forall y\, D(y)$ | $6_2[\alpha]$ |

Formulas 5 and 7 are complementary.

**Grading.** 6 points:

- 4 points for the steps 1–5
- 2 points for the steps 6–8

**Question 2** (16 points)**:** Recall that a relation $\to$ is called

*terminating* if there is no infinite descending chain $b_0 \to b_1 \to b_2 \to \cdots$;

*normalizing* if every $b \in A$ has a normal form;

*locally confluent* if $b \leftarrow a \to c$ implies there is a $d$ such that $b \to^* d \leftarrow^* c$;

*confluent* if $b \leftarrow^* a \to^* c$ implies there is a $d$ such that $b \to^* d \leftarrow^* c$.

Let

$$\{g(x, y, z) \to g(y, z, x),\ g(x, y, z) \to x\}$$

be a rewrite system over $\Sigma = (\{g/3,\ b/0,\ c/0,\ d/0\},\ \emptyset)$. Is the system (a) terminating? (b) normalizing? (c) locally confluent? (d) confluent? Briefly explain each of your answers.

**Answer.** (a) No, the system is not terminating. Consider the infinite chain $g(b, c, d) \to g(c, d, b) \to g(d, b, c) \to g(b, c, d) \to \cdots$.

(b) Yes, the system is normalizing. Every term has a normal form in which all occurrences of $g$ have been eliminated and are replaced by one of their arguments (in which $g$ has been recursively eliminated). For example, the normal form of $b$ is $b$, the normal forms of $g(b, c, d)$ are $b$, $c$, and $d$, and the normal forms of $g(g(b, c, b), c, b)$ are $b$ and $c$.

(c) Yes, the system is locally confluent. There is only one critical pair:

Between the first rule at position $\varepsilon$ and a renamed copy of the second rule:
mgu $\{x \mapsto x', y \mapsto y', z \mapsto z'\}$,
$g(y', z', x') \leftarrow g(x', y', z') \to x'$,
critical pair: $\langle g(y', z', x'), x' \rangle$.
The pair is joinable: $g(y', z', x') \to g(z', x', y') \to g(x', y', z') \to x'$.

(d) No, the system is not confluent. Consider the two chains $g(b, c, d) \to b$ and $g(b, c, d) \to g(c, d, b) \to c$. There is no way to join $b$ and $c$, which are in normal form.

**Grading.** 16 points:

- 4 points for each of the four parts
  - 2 points for "Yes" or "No"
  - 2 points for explanation

**Question 3** (14 points)**:** Apply the Knuth–Bendix procedure to the set of equations

$$g(s(x), y) \approx s(g(x, y)) \qquad (1)$$
$$g(x, s(y)) \approx s(g(x, y)) \qquad (2)$$

and transform it into a finite convergent term rewrite system. Use the Knuth–Bendix ordering with weight 1 for all symbols and variables and the precedence $g \succ s$. Document all the steps of the procedure.

**Answer.** We first apply "Orient" to (1) and (2), which results in the rewrite rules

$$g(s(x), y) \to s(g(x, y)) \qquad (3)$$
$$g(x, s(y)) \to s(g(x, y)) \qquad (4)$$

Then we apply "Deduce" between (3) and a renamed copy of (4), yielding the equation

$$s(g(s(x), y')) \approx s(g(x, s(y'))) \qquad (5)$$

We can use "Simplify-Eq" to simplify the left-hand side of the equation using (3):

$$s(s(g(x, y')) \approx s(g(x, s(y'))) \qquad (6)$$

We can use "Simplify-Eq" to simplify the right-hand side of the result using (4):

$$s(s(g(x, y')) \approx s(s(g(x, y'))) \qquad (7)$$

This is a trivial equation, which we can delete using "Delete."

At this point, no more critical pairs exist. Thus the set $\{(3), (4)\}$ is a finite convergent term rewrite system.

**Grading.** 14 points:

- 4 points for final result
- 4 points for correct orientation
- 3 points for deduction step
- 3 points for simplification and deletion

**Question 4** (18 points)**:** Let $\Sigma = (\{f/2,\ g/2,\ h/1,\ b/0\}, \emptyset)$. Consider the following set of equations:

$$\begin{aligned} f(b, y) &\approx y \\ f(h(x), y) &\approx h(f(x, y)) \\ g(b, y) &\approx b \\ g(h(x), y) &\approx f(y, g(x, y)) \end{aligned}$$

(a) Specify a precedence for the lpo so that the left-hand side of each equation in the set is greater than the corresponding right-hand side according to that instance of the lpo. There is no need to explain your answer.

**Answer.** $g \succ f \succ h \succ b$.

**Grading.** 8 points:

- 4 points subtracted if only three out of four equations are oriented correctly
- 0 points if only two out of four equations are oriented correctly

(b) Explain why the kbo cannot be used to orient all of these equations from left to right.

**Answer.** The fourth equation has two occurrences of $y$ on the right-hand side but only one occurrence on the left-hand side. This is forbidden by the kbo.

**Grading.** 2 points

(c) Specify a precedence, symbol weights, and a variable weight for the kbo so that the left-hand side of each of the *first three* equations is greater than the corresponding right-hand side according to that instance of the kbo. There is no need to explain your answer.

**Answer.** $g \succ f \succ h \succ b$ with weight 1 for all symbols and variables.

**Grading.** 8 points:

- 4 points subtracted if the weights are missing or illegal
- 0 points if only two out of three equations are oriented correctly

**Question 5** (16 points)**:** For this question, we use the lpo with the precedence $k \succ j \succ i \succ f \succ d \succ c \succ b$ as the term ordering.

(a) Fill in the table below by computing the rewrite systems $R_C$ and the increments $E_C$ for the set of ground clauses $N$ consisting of

$$i \approx b \vee j \approx c \qquad (1)$$
$$f(i) \not\approx f(b) \vee k \approx d \qquad (2)$$
$$i \approx b \qquad (3)$$
$$f(i) \not\approx f(b) \vee k \not\approx d \qquad (4)$$
$$i \not\approx b \vee j \approx c \qquad (5)$$

| Iter. | Clause $C$ | $R_C$ | $E_C$ |
|---|---|---|---|
| 0 | $i \approx b$ | $\emptyset$ | $\{i \to b\}$ |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |

**Answer.**

| Iter. | Clause $C$ | $R_C$ | $E_C$ |
|---|---|---|---|
| 0 | $i \approx b$ | $\emptyset$ | $\{i \to b\}$ |
| 1 | $i \approx b \vee j \approx c$ | $\{i \to b\}$ | $\emptyset$ |
| 2 | $i \not\approx b \vee j \approx c$ | $\{i \to b\}$ | $\{j \to c\}$ |
| 3 | $f(i) \not\approx f(b) \vee k \approx d$ | $\{i \to b, j \to c\}$ | $\{k \to d\}$ |
| 4 | $f(i) \not\approx f(b) \vee k \not\approx d$ | $\{i \to b, j \to c, k \to d\}$ | $\emptyset$ |

**Grading.** 12 points:

- 1 point for each of the twelve table entries, acknowledging Folgefehler

(b) Which is the smallest clause $C \in N$ such that $C$ is neither productive nor true in $R_C$?

**Answer.** The smallest such clause is $f(i) \not\approx f(b) \vee k \not\approx d$.

**Grading.** 2 points

- Acknowledge Folgefehler if the answer to part (a) has no such clause or a different clause

(c) Is $N$ saturated up to redundancy? Explain briefly.

**Answer.** No. By the model construction theorem, the existence of the clause in part (b) means that the set $N$ is not saturated up to redundancy.

**Grading.** 2 points

- Acknowledge Folgefehler if the answer to part (b) is that there exists no such clause.

**Question 6** (18 points)**:** Let us call a clause *grumpy* if it contains at least one negative literal.

(a) Prove that every inference of the resolution calculus from grumpy premises generates a grumpy conclusion.

**Answer.** We inspect the rules of the resolution calculus. For "Resolution," if the left premise is grumpy, then $D$ must contain a negative literal. Hence $D \vee C$ and its instance $(D \vee C)\sigma$ contain a negative literal. For "Factorization," if the premise is grumpy, then $C$ must contain a negative literal. Hence $D \vee C$ and its instance $(D \vee C)\sigma$ contain a negative literal.

**Grading.** 6 points:

- 3 points for "Resolution"
- 3 points for "Factorization"
- 0 points subtracted if $D \vee C$ instead of $(D \vee C)\sigma$

(b) Using the result stated in part (a) and the refutational completeness of resolution, prove that all sets of grumpy clauses are satisfiable. Hint: $\bot$ is not grumpy.

**Answer.** Let $N$ be a set of grumpy clauses. Consider the set $M$ defined inductively as the smallest set that includes $N$ and that is closed under the application of rules of the resolution calculus. Clearly, $M$ is saturated up to redundancy. Moreover, because inferences preserve grumpiness, every clause in $M$ is grumpy. As a result, the nongrumpy empty clause is not in $M$. By refutational completeness of resolution, $M$ is satisfiable, and hence $N \subseteq M$ is satisfiable.

**Grading.** 6 points:

- 3 points for the saturation process leading to $M$
- 3 points for observation that $\bot$, being not grumpy, is not in $M$

(c) Re-prove the result of part (b) using basic model theory. Specifically, for a given set $N$ of grumpy clauses, exhibit an algebra $\mathcal{A}$ such that $\mathcal{A} \models N$.

**Answer.** (c) Consider the algebra $\mathcal{A}$ that interprets every predicate symbol as $\emptyset$, thereby returning 0 regardless of the arguments. Since each clause in $N$ contains a negative literal, that literal will be true in $\mathcal{A}$, and hence each clause in $N$ will be true in $\mathcal{A}$.

**Grading.** 6 points:

- 4 points for definition of $\mathcal{A}$
- 2 points for explanation why $\mathcal{A} \models N$