Regular Examination in the Course Automated Theorem Proving

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

Winter Term 2025/26
20.02.2026

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:

Hereby, I confirm the correctness of the above information:

Signature

With your signature, you confirm that you are sufficiently healthy at the beginning of the examination and that you accept the examination bindingly.

Please leave the following table blank:

Question	1	2	3	4	5	6	$\sum$
Points	17	21	16	16	12	18	100
Score							

Instructions

You have **120 minutes** at your disposal. Written or electronic aids are not permitted except for normal watches. Carrying forbidden devices, even turned off, will be considered a cheating attempt.

Write your full name and matriculation number clearly legible on this 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 3–13**. 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 such as this one.

Grading. And blocks such as this one specify the grading scheme.

Question 1 (17 points)**:** (a) We call a clause *depressed* if it contains at least as many negative literals as it contains positive literals. For example, $P(a) \vee \neg Q(b) \vee \neg R$ and $P(a) \vee \neg Q(b) \vee \neg R \vee S$ are depressed. Prove or disprove that every inference of the resolution calculus from depressed premises generates a depressed conclusion.

Answer. We present a proof.

For "Resolution," suppose that the left premise has p_1 positive and $n_1 \geq p_1$ negative literals, and the right premise has p_2 positive and $n_2 \geq p_2$ negative literals. The conclusion has $p_1 - 1 + p_2$ positive literals (with $p_1 - 1$ because of the resolved positive literal from the left premise) and $n_1 + n_2 - 1$ negative literals (with $n_2 - 1$ because of the resolved negative literal from the right premise). Since $n_1 \geq p_1$ and $n_2 \geq p_2$, we have $n_1 + n_2 - 1 \geq p_1 - 1 - p_2$, and the conclusion is depressed.

For "Factorization," suppose that the premise has p positive and $n \geq p$ negative literals. The conclusion has $p - 1$ positive literals and n negative literals (since two identical positive literals are merged). Thus, we have $n \geq p - 1$, and the conclusion is depressed.

Grading. 7 points:

- 1 point for prove
- 3 points for "Resolution"
- 3 points for "Factorization"

(b) We call a clause *balanced* if it contains exactly as many negative literals as it contains positive literals. Prove or disprove that every inference of the resolution calculus from balanced premises generates a balanced conclusion.

Answer. There exists a counterexample: The premise $P \vee P \vee \neg Q \vee \neg R$ is balanced, but the conclusion $P \vee \neg Q \vee \neg R$ of a "Factoring" inference with this premise is not balanced.

Grading. 5 points:

- 1 point for disprove
- 4 points for counterexample (or abstract counterargument)

(c) We call a clause *unbalanced* if it is not balanced (as per subquestion (b)). Prove or disprove that every inference of the resolution calculus from unbalanced premises generates an unbalanced conclusion.

Answer. There exists a counterexample: The premises P and $\neg P$ are unbalanced, but the conclusion $\perp$ of a "Resolution" inference with these two premises is balanced (with 0 positive and 0 negative literals).

Grading. 5 points:

- 1 point for disprove
- 4 points for counterexample (or abstract counterargument)

Question 2 (21 points)**:** Consider the following entailment between two propositional formulas:

$$P \rightarrow (Q \vee R) \models (P \wedge \neg Q) \rightarrow R$$

(a) Express the entailment as a set of clauses by negating the conjecture and using the CNF transformation. Document each step.

Answer. Negating the conjecture and conjoining:

$$(P \rightarrow (Q \vee R)) \wedge \neg((P \wedge \neg Q) \rightarrow R)$$

CNF transformation:

$$\begin{array}{rl}
 & (P \rightarrow (Q \vee R)) \wedge \neg((P \wedge \neg Q) \rightarrow R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge \neg((P \wedge \neg Q) \rightarrow R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge \neg(\neg(P \wedge \neg Q) \vee R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge \neg((\neg P \vee \neg\neg Q) \vee R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge \neg((\neg P \vee Q) \vee R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge (\neg(\neg P \vee Q) \wedge \neg R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge ((\neg\neg P \wedge \neg Q) \wedge \neg R) \\
\Rightarrow_{CNF} & (\neg P \vee Q \vee R) \wedge ((P \wedge \neg Q) \wedge \neg R) \\
\text{i.e.} & (\neg P \vee Q \vee R) \wedge P \wedge \neg Q \wedge \neg R
\end{array}$$

Clauses:

$$\begin{gathered}
\neg P \vee Q \vee R \\
P \\
\neg Q \\
\neg R
\end{gathered}$$

Grading. 8 points:

- 2 points for the result of clausification
- 6 points if the six intermediate steps are present and correct

(b) Let $N = \{\neg S \vee T \vee U,\ S,\ \neg T,\ \neg U\}$ be a clause set. Use the DPLL procedure to show whether N is satisfiable or unsatisfiable. Document each DPLL step.

Answer. Initially, we set $M := \emptyset$. Since S contains the unit literal S, we set $M := \{S\}$. Since $\neg T$ contains the unit literal $\neg T$, we set $M := \{S, \neg T\}$. Since $\neg U$ contains the unit literal $\neg U$, we set $M := \{S, \neg T, \neg U\}$. At this point, $\neg S \vee T \vee U$ is false in M, and there is nowhere to backtrack to, so the clause set N is unsatisfiable.

Grading. 4 points:

- 1 point for initial state
- 1 point for each of the three unit propagations

(c) Use the (unordered) resolution calculus to derive the empty clause from the clause set $N = \{\neg S \vee T \vee U, S, \neg T, \neg U\}$. Document all of your steps.

Answer. From $\neg S \vee T \vee U$ and S, we obtain via "Resolution" $T \vee U$. From $T \vee U$ and $\neg T$, we obtain via "Resolution" U. From U and $\neg U$, we obtain via "Resolution" the empty clause.

Grading. 3 points:

- 1 point for each "Resolution" step

(d) Use the ground tableau calculus to prove the entailment

$$P \rightarrow (Q \vee R) \models (P \wedge \neg Q) \rightarrow R$$

You will first need to transform the entailment into a formula to refute. Use exactly the expansion rules seen in the lecture; do not preclausify the problem. Document all of your steps.

Answer. The formula to refute is $(P \rightarrow (Q \vee R)) \wedge \neg((P \wedge \neg Q) \rightarrow R)$.

1.	$(P \rightarrow (Q \vee R)) \wedge \neg((P \wedge \neg Q) \rightarrow R)$	[formula to refute]
2.	$P \rightarrow (Q \vee R)$	$1_1[\alpha]$
3.	$\neg((P \wedge \neg Q) \rightarrow R)$	$1_2[\alpha]$
4.	$(P \wedge \neg Q)$	$3_1[\alpha]$
5.	$\neg R$	$3_2[\alpha]$
6.	P	$4_1[\alpha]$
7.	$\neg Q$	$4_2[\alpha]$

- 9. $Q \vee R$ $2_2[\beta]$
  - 10. Q $8_1[\beta]$
  - 11. R $8_2[\beta]$
- 8. $\neg P$ $2_1[\beta]$

There are three paths, each of them closed: 7 and 10 are complementary along the first path, 5 and 11 are complementary along the second path, and 6 and 8 are complementary along the third path.

Grading. 6 points

- 1 point for the formula to refute
- 0.5 points (rounded up) for each correct step 2–11

Question 3 (16 points)**:** Let $\Sigma = (\{append/2,\ cons/2,\ nil/0,\ rev/1\},\ \emptyset)$, and let x, xs, ys be variables. Consider the following set of equations:

$$
\begin{aligned}
rev(nil) &\approx nil \\
rev(cons(x, xs)) &\approx append(rev(xs), cons(x, nil)) \\
append(nil, ys) &\approx ys \\
append(cons(x, xs), ys) &\approx cons(x, append(xs, ys))
\end{aligned}
$$

(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 explain your answer.

Answer. $rev \succ append \succ cons \succ nil$.

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) Note that the kbo cannot be used to orient these equations from left to right. Regardless of the weights chosen, the right-hand side of the second equation has a greater weight than the left-hand side. So let us ignore the second equation.

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

Answer. $rev \succ append \succ cons \succ nil$ 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 4 (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

$$\{f(x, y) \to f(y, x),\ f(x, y) \to d\}$$

be a rewrite system over $\Sigma = (\{f/2,\ b/0,\ c/0,\ d/0\},\ \emptyset)$. 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 $f(b, c) \to f(c, b) \to f(b, c) \to f(c, b) \to \cdots$.

(b) Yes, the system is normalizing. Every term t has a normal form, which is t itself if t is b, c, or d and d otherwise. For example, the normal form of b is b, the normal form of $f(b, c)$ is d, and the normal form of $f(f(b, c), b)$ is d.

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

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

(d) Yes, the system is confluent. In the definition of confluence, if a is the function symbol b, c, or d, then we can take the terms b, c, and d to be a as well. Otherwise, we can let the term d be the function symbol d.

Grading. 16 points:

- 4 points for each of the four parts
  - 1 point for "Yes" or "No"
  - 3 points for explanation

Question 5 (12 points)**:** Refute the following set of equational clauses using the superposition calculus:

$$pi \not\approx zero \qquad (1)$$

$$x \approx zero \vee div(one, x) \approx inv(x) \qquad (2)$$

$$abs(div(one, pi)) \not\approx abs(inv(pi)) \qquad (3)$$

Use the kbo with the precedence $abs \succ div \succ inv \succ one \succ pi \succ zero$ and weight 1 for all symbols and variables. Document all of your steps.

Answer. From (2) and (3), we obtain via "Negative Superposition"

$$pi \approx zero \vee abs(inv(pi)) \not\approx abs(inv(pi)) \qquad (4)$$

From (4), we obtain via "Equality Factoring"

$$pi \approx zero \qquad (5)$$

From (5) and (1), we obtain via "Negative Superposition"

$$zero \not\approx zero \qquad (6)$$

From (6), we obtain via "Equality Factoring" the empty clause.

Grading. 12 points:

- 3 points for each of the four inferences

Question 6 (18 points)**:** For this question, we use the signature $\Sigma = (\{f/1,\ g/1,\ b/0\}, \emptyset)$, the clause set $N = \{g(x) \approx b\}$ over Σ, and the kbo with $g \succ f \succ b$ and weights 1 for all symbols and variables.

(a) Complete the following table summarizing the first six iterations of the candidate interpretation construction. Hint: Recall that the clauses considered in the second column must be groundings of clauses in N.

Iter.	Clause C	R_C	E_C
0	$g(b) \approx b$	$\emptyset$	$\{g(b) \to b\}$
1			
2			
3	$g(f(f(b))) \approx b$		
4			
5			

Answer.

Iter.	Clause C	R_C	E_C
0	$g(b) \approx b$	$\emptyset$	$\{g(b) \to b\}$
1	$g(f(b)) \approx b$	$\{g(b) \to b\}$	$\{g(f(b)) \to b\}$
2	$g(g(b)) \approx b$	$\{g(b) \to b,\ g(f(b)) \to b\}$	$\emptyset$
3	$g(f(f(b))) \approx b$	$\{g(b) \to b,\ g(f(b)) \to b\}$	$\{g(f(f(b))) \to b\}$
4	$g(f(g(b))) \approx b$	$\{g(b) \to b,\ g(f(b)) \to b,\ g(f(f(b))) \to b\}$	$\emptyset$
5	$g(g(f(b))) \approx b$	$\{g(b) \to b,\ g(f(b)) \to b,\ g(f(f(b))) \to b\}$	$\emptyset$

Grading. 14 points:

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

(b) Give a model of N.

Answer. Consider the algebra $\mathcal{A}$ that interprets g as the constant function that maps every value to the interpretation of b. The interpretation of f is arbitrary.

Grading. 4 points