First Examination in the Course Automated Theorem Proving

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

Winter Term 2024/25
11 February 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	14	24	16	16	14	16	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 (14 points)**:** (a) A clause is called *Horn* if it contains at most one positive literal. Prove that every inference of the resolution calculus from Horn premises generates a Horn conclusion.

Answer. For "Resolution," since B is the positive literal in the first premise, D consists exclusively of negative literals, and C contains at most one positive literal. Hence $D \vee C$ and its instance $(D \vee C)\sigma$ contain at most one positive literal. The case of "Factorization" is impossible: The premise of that rule is never Horn.

Grading. 5 points:

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

(b) A clause is called *dual-Horn* if it contains at most one negative literal. Prove that every inference of the resolution calculus from dual-Horn premises generates a dual-Horn conclusion.

Answer. For "Resolution," since $\neg A$ is the negative literal in the second premise, C consists exclusively of positive literals, and D contains at most one negative literal. Hence $D \vee C$ and its instance $(D \vee C)\sigma$ contain at most one negative literal. For "Factorization," since A and B are positive, C is allowed to contain at most one negative literal. This means $C \vee A$ and its instance $(C \vee A)\sigma$ contain at most one negative literal.

Grading. 5 points:

- 3 points for "Resolution"
- 2 points for "Factorization"

(c) Let N be a clause set consisting exclusively of Horn and dual-Horn clauses. Does $\mathit{Res}(N)$ also consist exclusively of Horn and dual-Horn clauses? Explain your answer.

Answer. No, it does not. A counterexample follows. We take $N := \{P \vee P \vee P, \neg P \vee \neg P \vee \neg P\}$. The first clause is dual-Horn; the second clause is Horn. If we resolve the two clauses together, we obtain $P \vee P \vee \neg P \vee \neg P$, which is neither Horn nor dual-Horn.

Grading. 4 points:

- 2 points for "No"
- 2 points for counterexample (or abstract proof)

Question 2 (24 points)**:** Consider the following entailment between two propositional formulas:

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

(a) Use the DPLL procedure to prove the entailment. You will first need to express the entailment as a set of clauses using the CNF transformation. Document each CNF and DPLL step.

Answer. We start with the CNF transformation:

$$\begin{aligned}
& ((P \wedge Q) \rightarrow R) \wedge \neg((P \rightarrow R) \vee (Q \rightarrow R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge \neg((P \rightarrow R) \vee (Q \rightarrow R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge \neg((\neg P \vee R) \vee (Q \rightarrow R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge \neg((\neg P \vee R) \vee (\neg Q \vee R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge (\neg(\neg P \vee R) \wedge \neg(\neg Q \vee R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge ((\neg\neg P \wedge \neg R) \wedge \neg(\neg Q \vee R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge ((P \wedge \neg R) \wedge \neg(\neg Q \vee R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge ((P \wedge \neg R) \wedge (\neg\neg Q \wedge \neg R)) \\
\Rightarrow_{CNF}\ & (\neg(P \wedge Q) \vee R) \wedge ((P \wedge \neg R) \wedge (Q \wedge \neg R)) \\
\Rightarrow_{CNF}\ & ((\neg P \vee \neg Q) \vee R) \wedge ((P \wedge \neg R) \wedge (Q \wedge \neg R))
\end{aligned}$$

Exploiting the associativity and commutativity of $\wedge$ and $\vee$, we obtain

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

Thus $N = \{\neg P \vee \neg Q \vee R,\ P,\ Q,\ \neg R\}$.

Next, we invoke the DPLL procedure. Initially, we set $M := \emptyset$. Since P contains the unit literal P, we set $M := \{P\}$. Since Q contains the unit literal Q, we set $M := \{P, Q\}$. Since $\neg R$ contains the unit literal $\neg R$, we set $M := \{P, Q, \neg R\}$. At this point, $\neg P \vee \neg Q \vee R$ is false in M, and there is nowhere to backtrack to, so the clause set N is unsatisfiable.

Grading. 14 points:

- 10 points for clausification
  - 4 points for the result of clausification
  - 6 points if at least six intermediate steps are present and correct
- 4 points for the DPLL procedure
  - 1 point for initial state
  - 1 point for each of the three unit propagations
  - Folgefehler are acknowledged

(b) Use the (unordered) resolution calculus to derive the empty clause from the clause set $N = \{\neg P \vee \neg Q \vee R,\ P,\ Q,\ \neg R\}$. Document all of your steps.

Answer. From $\neg P \vee \neg Q \vee R$ and P, we obtain via "Resolution" $\neg Q \vee R$. From $\neg Q \vee R$ and Q, we obtain via "Resolution" R. From R and $\neg R$, we obtain via "Resolution" the empty clause.

Grading. 4 points

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

$$(P \land Q) \to R \models (P \to R) \lor (Q \to R)$$

You will first need to transform the entailment into a formula to refute. Use exactly the expansion rules seen in the lecture. Document all of your steps.

Answer. The formula to refute is $((P \land Q) \to R) \land \neg((P \to R) \lor (Q \to R))$.

1.	$(P \land Q \to R) \land \neg((P \to R) \lor (Q \to R))$	[formula to refute]
2.	$P \land Q \to R$	$1_1[\alpha]$
3.	$\neg((P \to R) \lor (Q \to R))$	$1_2[\alpha]$
4.	$\neg(P \to R)$	$3_1[\alpha]$
5.	$\neg(Q \to R)$	$3_2[\alpha]$
6.	P	$4_1[\alpha]$
7.	$\neg R$	$4_2[\alpha]$
8.	Q	$5_1[\alpha]$
9.	$\neg R$	$5_2[\alpha]$

$$\begin{array}{ccc}
 & 9. \quad \neg R & \\
 \swarrow & & \searrow \\
10. \quad \neg(P \land Q) \quad 2_1[\beta] & & 13. \quad R \quad 2_2[\beta]
\end{array}$$

$$\begin{array}{ccc}
 & 10. \quad \neg(P \land Q) & \\
 \swarrow & & \searrow \\
11. \quad \neg P \quad 10_1[\beta] & & 12. \quad \neg Q \quad 10_2[\beta]
\end{array}$$

There are three paths, each of them closed: 6 and 11 are complementary along the first path; 8 and 12 are complementary along the second path; and 7 and 13 are complementary along the third path.

Grading. 6 points:

- 0.5 points (rounded up) for each correct step

Question 3 (16 points)**:** Let $\Sigma = (\{f/1,\ g/2,\ h/2,\ b/0\}, \emptyset)$. Consider the following set of equations:

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

(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. $f \succ h \succ g \succ b$.

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

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. $f \succ h \succ g \succ b$ 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 4 (16 points)**:** Let $\Sigma = (\Omega, \emptyset)$ with $\Omega = \{b/0,\ c/0,\ f/2,\ g/1,\ h/2\}$ and let R be the following rewrite system:

$$\begin{aligned} h(y,c) &\to g(y) && (1) \\ f(x,x) &\to g(h(x,b)) && (2) \\ h(h(x,b),x) &\to f(x,c) && (3) \end{aligned}$$

(a) Find a kbo $\succ$ such that $\to_R \subseteq \succ$. Specify the weights and precedence of the ordering. There is no need to explain your answer.

Answer. There are many kbo instances $\succ$ such that $\to_R \subseteq \succ$. One possibility: $w(f) = 5$, $w(h) = 3$, $w(g) = w(b) = w(c) = w(x) = 1$. In this case, the precedence does not matter.

Grading. 6 points:

- 3 points subtracted if the weights are missing or illegal
- 3 points subtracted if only two out of three rewrite rules are oriented correctly

(b) Compute all critical pairs between rules in R and determine whether they are joinable in R. There should be two critical pairs.

Answer. Between (3) at position 1 and a renamed copy of (3):
mgu $\{x \mapsto h(b,b),\ x' \mapsto b\}$,
$f(h(b,b),c) \leftarrow h(h(h(b,b),b),h(b,b)) \to h(f(b,c),h(b,b))$,
critical pair: $\langle f(h(b,b),c),\ h(f(b,c),h(b,b))\rangle$.
Both terms are in normal form, therefore not joinable.

Between (3) at position ε and (1):
mgu $\{x \mapsto c,\ y \mapsto h(c,b)\}$,
$f(c,c) \leftarrow h(h(c,b),c) \to g(h(c,b))$,
critical pair: $\langle f(c,c),\ g(h(c,b))\rangle$.
$f(c,c)$ can be rewritten to $g(h(c,b))$ using (3), therefore joinable.

Grading. 10 points:

- 4 points for each of the two critical pairs
  - 2 points subtracted if variables are not renamed
- 1 point for each "joinable"/"nonjoinable" answer

Question 5 (14 points)**:** Refute the following set of equational clauses using the superposition calculus:

$$\begin{aligned} f(x) \approx f(c) \lor f(x) \approx f(b) \quad &(1) \\ f(d) \not\approx f(b) \quad &(2) \\ f(d) \not\approx f(c) \quad &(3) \end{aligned}$$

Use the lpo with the precedence $f \succ d \succ c \succ b$. Document all of your steps.

Answer. From (1) and (3), we obtain via "Negative Superposition"

$$f(d) \approx f(b) \lor f(c) \not\approx f(c) \quad (4)$$

From (4) and (2), we obtain via "Negative Superposition"

$$f(c) \not\approx f(c) \lor f(b) \not\approx f(b) \quad (5)$$

From (5), we obtain via "Equality Factoring"

$$f(b) \not\approx f(b) \quad (6)$$

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

Grading. 14 points:

- 4 points for each of the two "Negative Superposition" inferences
- 3 points for each of the two "Equality Factoring" inferences
- 2 points subtracted for each extra wrong inference

Question 6 (16 points)**:** For this question, we use the signature $\Sigma = (\{f/1,\ b/0,\ c/0\}, \emptyset)$, the clause set $N = \{f(x) \approx b\}$ over Σ, and the kbo with $f \succ c \succ b$ and weights 1 for all symbols and variables.

(a) Complete the following table summarizing the first five 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	$f(b) \approx b$		$\{f(b) \to b\}$
1			
2			
3	$f(f(c)) \approx b$		
4			

Answer.

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

Grading. 12 points:

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

(b) Determine R_∞ for the clause set N.

Answer. Extrapolating from the answer to part (a), we conclude that only the first two ground clauses are productive. Thus $R_\infty = \{f(b) \to b,\ f(c) \to b\}$.

Grading. 4 points

- acknowledge Folgefehler