Some Fragments Towards Establishing Completeness Properties of Second-Order Quantifier Elimination Methods

Christoph Wernhard, TU Dresden

Deduktionstreffen 2015

1. Background: Second-Order Quantifier Elimination

Input: a formula **with second-order quantifiers**

$$\exists p \forall x\, (\neg qx \lor px) \land (\neg px \lor rx)$$

Output: an **equivalent first-order formula** without new symbols

$$\forall x\, \neg qx \lor rx$$

- Variants: uniform interpolation, forgetting, projection
- Applications are:
  - ontology reuse and analysis, information hiding [13, 21, 20]
  - circumscription [10, 29]
  - abduction in logic programming [30]
  - and many more [11]

2. The Quest for Completeness of Elimination Methods

- Which **classes of formulas** ensure that there is an elimination result?
- Which **methods** guarantee success on particular such classes?

3. Elimination Methods: Basic Approaches

- **Direct approach**, **Ackermann approach**, **DLS** [27, 10, 6, 25, 2, 1]
  Rewrite $\exists p\, F$ until all subformulas $\exists p\, F'$ match a form for which the elimination result is known (*"Ackermann's Lemma"*)
- **Resolvent generation**, **SCAN** [12, 1]
  Convert $\exists p\, F$ to CNF and replace all clauses with p by their resolvents upon p

4. Some Known Completeness Properties for Elimination

- Elimination on **propositional** formulas succeeds
$$\exists p\, F[p] \equiv F[\top] \lor F[\bot] \quad (1)$$
- A variant of the direct approach succeeds on **relational monadic formulas** (the Löwenheim class) [19, 26, 2]
  Successive elimination of all predicates then provides a **decision procedure**
- Sahlqvist formulas are modal formulas with first-order correspondence properties that can be computed with the *Sahlqvist-van Benthem method* [24, 3, 4]
  Completeness for Sahlqvist formulas has been shown for SCAN [14] and DLS [5]
- The **success of DLS has been characterized syntactically** [5]
- There are specialized methods for **modal and description logics**, e.g., [16, 17, 28, 18, 20, 6, 25, 31]

5. Ackermann's Lemma [1]

- **Notation:** If $F[p]$ is a first-order formula (possibly with occurrences of p) and $G[x]$ is a first-order formula (possibly with free occurrences of x) without occurrences of p and of variables bound in $F[p]$, then $F[G]$ **denotes** $F[p]$ **with all occurrences** $p(t_i)$ **of** p **replaced by** $G[t_i]$, that is, by $G[x]$ with all free occurrences of x replaced by t_i
- **Ackermann's Lemma:** If p is not in $A[x]$ and only **positive** in $B[p]$, then
$$\exists p\, (\forall x\, \neg px \lor A[x]) \land B[p] \;\equiv\; B[A] \quad (2)$$
  Analogously, if p is not in $A[x]$ and only negative in $B[p]$, then
$$\exists p\, (\forall x\, px \lor A[x]) \land B[p] \;\equiv\; B[\neg A] \quad (3)$$
- **Example:**
$$\exists p\, (\forall x\, \neg px \lor qx) \land ((\exists y\, py \land ry) \lor pa) \;\equiv\; (\exists y\, qy \land ry) \lor qa$$

6. Outline of DLS [27, 10, 15, 5]

1. **Preprocessing:** Convert $\exists p\, F$ to an equivalent formula of the form
$$\exists x_1 \ldots \exists x_k \exists p\, (A_1 \land B_1) \lor \ldots \lor (A_n \land B_n), \quad (4)$$
   where p is only negative in the A_i and only positive in the B_i
   - 1.1 Convert to negation normal form
   - 1.2 Move quantifiers inward/outward:
     $Qx\, F[x] \otimes G \equiv (Qx\, F[x]) \otimes G$, if x not in G
   - 1.3 Distribute $\land$ over $\lor$:
     $F \land (G \lor H) \Rightarrow (F \land G) \lor (F \land H)$,
     if p occurs positively as well as negatively in $G \land H$

   This step might fail

   Convert (4) to
$$\exists x_1 \ldots \exists x_k (\exists p\, A_1 \land B_1) \lor \ldots \lor (\exists p\, A_n \land B_n), \quad (5)$$
   and process each $\exists p\, A_i \land B_i$ individually
2. **Preparation for Ackermann's Lemma:** Convert
$$\exists p\, A \land B, \quad (6)$$
   where p is only negative in A and only positive in B to
$$\exists f_1 \ldots \exists f_m \exists p\, (\forall x\, px \lor A'[x]) \land B'[p], \quad (7)$$
   where the f_i are fresh Skolem functions, p is not in A' and is only positive in $B'[p]$

   This is always possible, also with the roles of A and B switched
3. **Application of Ackermann's Lemma:** Ackermann's Lemma applied to the subformula of (7) starting at $\exists p$ yields
$$\exists f_1 \ldots \exists f_m\, B'[A'] \quad (8)$$
   Un-Skolemize, which might fail
4. **Simplification**

7. Examples where DLS Fails Unnecessarily

- A **monadic** formula that requires distribution of $\lor$ over $\land$ [2, 31]:
$$\begin{aligned} & \exists p\, \forall x\, (px \land qx) \lor (\neg px \land rx) \\ \equiv\; & \exists p\, \forall x\, (px \lor rx) \land (qx \lor \neg px) && \text{Distribute } \lor \text{ over } \land \\ \equiv\; & \exists p\, (\forall x\, px \lor rx) \land (\forall x\, qx \lor \neg px) && \text{Move } \forall \text{ inward} \end{aligned} \quad (9)$$
- **"Reasoning"** is required to obtain the **"p-separated"** form (6):
$$\begin{aligned} & \exists p\, (\forall x \forall y\, \neg px \lor \neg qxy \lor (py \land ry)) \;\land\; \forall x\, \neg rx \\ \equiv\; & \exists p\, (\forall x \forall y\, \neg px \lor \neg qxy \lor (py \land \bot)) \;\land\; \forall x\, \neg rx \\ \equiv\; & \exists p\, (\forall x \forall y\, \neg px \lor \neg qxy) \;\land\; \forall x\, \neg rx \\ \equiv\; & \exists p\, \top \land ((\forall x \forall y\, \neg px \lor \neg qxy) \;\land\; \forall x\, \neg rx) \end{aligned} \quad (10)$$

8. Determining Separability and Separation Formulas

Theorem 1. Let F be a first-order formula and p a predicate. Then:

(i) A first-order formula can be constructed that is valid if and only if there are first-order formulas A and B (without symbols not in F) where p is only negative in A and only positive in B such that $F \equiv A \land B$

(ii) In case formulas A, B according to (i) exist, the pairs of formulas A, B meeting the conditions of (i) can be characterized exactly as the Craig/Lyndon interpolants of first-order formulas constructed in a specific way

- The theorem says that:
  - ***p*-separability can be reduced to first-order validity**
  - **In case of *p*-separability, *p*-separations $A \land B$ can be computed by Craig/Lyndon interpolation [7, 8, 9, 22]**
  - **Moreover, all *p*-separations (modulo equivalence) are such Craig/Lyndon interpolants**
- The theorem can be generalized, e.g. that A may only contain predicates from a given set. If these are all monadic, this ensures success of subsequent steps of the elimination [31]

Is an analog theorem for form (4) instead of (6) possible?

9. Proof Sketch of Theorem 1

- We use notation for quantification upon "a predicate *in a polarity*" ("literal forgetting"). For fresh q:
$$\begin{aligned} \exists{+}p\, F[p] &\text{ stands for } \exists q\, F[q] \land (\forall x\, \neg qx \rightarrow \neg px) \\ \exists{-}p\, F[p] &\text{ stands for } \exists q\, F[q] \land (\forall x\, qx \rightarrow px) \end{aligned} \quad (11)$$
- The semantic conditions on A, B can be expressed as:
$$A \land B \models F,\;\; F \models A \land B,\;\; A \equiv \exists{+}p\, A,\;\; B \equiv \exists{-}p\, B \quad (12)$$
- Note: Formulas that are equivalent to A, B and also do not *syntactically* contain p negatively and positively, resp., can be obtained from A, B by Craig/Lyndon interpolation, see [23, Introduction]
- From (12) follows $\exists{+}p\, F \models A$ and $\exists{-}p\, F \models B$ and thus
$$(\exists{+}p\, F) \land (\exists{-}p\, F) \models F \quad (13)$$
  Let $F[p] = F$. Then (13) holds iff, for fresh q, r:
$$F[q] \land (\forall x\, \neg qx \rightarrow \neg px) \land F[r] \land (\forall x\, rx \rightarrow px) \models F[p] \quad (14)$$
  Given (14), we construct A as Craig/Lyndon interpolant:
$$F[q] \land (\forall x\, \neg qx \rightarrow \neg px) \models A \models (F[r] \land (\forall x\, rx \rightarrow px)) \rightarrow F[p] \quad (15)$$
  and then B as Craig/Lyndon interpolant:
$$F[r] \land (\forall x\, rx \rightarrow px) \models B \models A \rightarrow F[p] \quad (16)$$
- It can further be shown that if (14) holds, then all A, B satisfying (15) and (16) also satisfy (12)
- A related generalization of Craig interpolation is [8, Lemma 2]

10. Eliminability by Uniform Replacement

Definition 1. A predicate p is **eliminable by uniform replacement** from a first-order formula $F[p]$ (briefly $F[p]$ is **EBUR**) if and only if there is a first-order formula G without occurrences of p and of variables bound in $F[p]$ such that
$$\exists p\, F[p] \equiv F[G]$$

- EBUR formulas have some "good" properties:
  - **P1. Determining whether for given $F[p]$ and G it holds that $\exists F[p] \equiv F[G]$ can be reduced to first-order validity**
  - **P2. The $F[p]$ that are EBUR are recursively enumerable**
  - **P3. For a given EBUR $F[p]$, the G such that $\exists p\, F[p] \equiv F[G]$ are recursively enumerable**
  - **P4. For a given EBUR $F[p]$, a first-order formula F' whose symbols are all from $F[p]$ (and which does not contain p) such that $\exists F[p] \equiv F'$ can be computed**
- EBUR formulas cover some well-known cases of successful elimination:
  - If $\exists p\, F[p]$ matches the left side of **Ackermann's Lemma**, then $F[p]$ is EBUR
  - Propositional $F[p]$ are EBUR
  - If there is a G such that $F[p] \models \forall x\, px \leftrightarrow G$, then $F[p]$ is EBUR

11. Proof Sketches of the Properties of EBUR formulas

- P1: Recall that p does not occur in $F[G]$
$$\begin{aligned} \exists p\, F[p] \models F[G] &\text{ holds if and only if } F[p] \models F[G] \quad (17) \\ F[G] \models \exists p\, F[p] &\text{ holds in general} \quad (18) \end{aligned}$$
- P2 and P3 follow from P1
- P4: F' can be obtained by Craig interpolation
$$F[p] \models F' \models F[G] \quad (19)$$

12. Uniform Replacement: Justification of Covered Cases

- Ackermann's Lemma:
$$\begin{aligned} \exists p\, F[p] &\equiv \exists p\, (\forall x\, \neg p(x) \lor A) \land B[p] \equiv B[A] \\ &\equiv \quad\; (\forall x\, \neg A \lor A) \quad\; \land B[A] \equiv F[A] \end{aligned} \quad (20)$$
- Propositional formulas: $F[p]$ can be brought into the form
$$(\neg p \lor A) \land (p \lor B) \land C, \quad (21)$$
  with p not in A, B, C, which matches Ackermann's Lemma
- Entailed definition: $\exists p\, F[p] \;\equiv\; \exists p\, F[p] \land (\forall x\, px \leftrightarrow G) \;\equiv\; F[G]$
  Alternatively, a match with Ackermann's Lemma can be established:
$$F[p] \;\equiv\; F[G] \land (\forall x\, px \rightarrow G) \land (\forall x\, px \leftarrow G) \quad (22)$$
- In the following example $F[p]$ is EBUR but does not match Ackermann's Lemma; after rewriting, a **subformula** matches it:
$$\begin{aligned} \exists p\, F[p] &\equiv \exists p \exists y\, (\forall x\, qxy \rightarrow px) \land (\forall x\, px \rightarrow rxy) \\ &\equiv \exists y \exists p\, (\forall x\, qxy \rightarrow px) \land (\forall x\, px \rightarrow rxy) \end{aligned} \quad (23)$$

13. Uniform Replacement: Questions and Observations

- Is there a **general method** to compute solutions G that is better than naive generating and validity testing?
- Is consideration of G that only use symbols from $F[\,]$ sufficient?
- If $F[p]$ is EBUR, how far are equivalent formulas also EBUR?

- Ways to **characterize** $\exists p\, F[p] \equiv F[G]$:
  - With the second-order analog to $p(t) \equiv \forall x\, p(x) \lor x \neq t$:
$$\exists p\, F[p] \models \forall p\, F[p] \lor \neg(\forall x\, px \leftrightarrow G) \quad (24)$$
  - G is a "counter-definiens" of p (q is fresh):
$$F[q] \land \neg F[p] \models \neg(\forall x\, px \leftrightarrow G) \quad (25)$$
  - G is a "counter-definiens" of p – expressed differently:
$$F[q] \land \neg F[p] \models (\exists x\, \neg px \leftrightarrow G) \quad (26)$$
    If $\exists$ would be replaced there by $\forall$, we could obtain G as definiens of $\neg p$ by Craig interpolation
  - Entailment of a disjunction of related definitions for some $k \geq 0$ and ground terms $a_1, \ldots, a_k$ (by Herbrand's Theorem):
$$F[q] \land \neg F[p] \models (\neg pa_1 \leftrightarrow G[a_1]) \lor \ldots \lor (\neg pa_k \leftrightarrow G[a_k]) \quad (27)$$

14. A Tentative Generalization of Uniform Replacement

Definition 2. A predicate p is **eliminable by multiform replacement** from a first-order formula $F[p]$ if and only if there are a number $n \geq 0$, a first-order formula $F'[p, \ldots, p]$ and first-order formulas $G_1, \ldots, G_n$ without occurrences of p and of variables bound in F' such that $F[p] \equiv F'[p, \ldots, p]$ and
$$\exists p\, F'[p, \ldots, p] \;\equiv\; \exists p_1 \ldots \exists p_n\, F'[p_1, \ldots, p_n] \;\equiv\; F'[G_1, \ldots, G_n]$$
(The arity of $[p, \ldots, p]$ is n, each position representing a partition of the occurrences of p. The p_i are fresh)

- Example:
$$\begin{aligned} & \exists p\, ((\forall x\, px \lor qx) \land (\forall x\, \neg px \lor rx)) \lor ((\forall x\, px \lor sx) \land (\forall x\, \neg px \lor tx)) \\ \equiv\; & \exists p_1 \exists p_2\, ((\forall x\, p_1x \lor qx) \land (\forall x\, \neg p_1x \lor rx)) \lor \\ & \qquad ((\forall x\, p_2x \lor sx) \land (\forall x\, \neg p_2x \lor tx)) \\ \equiv\; & ((\forall x\, qx \lor qx) \land (\forall x\, \neg qx \lor rx)) \lor ((\forall x\, sx \lor sx) \land (\forall x\, \neg sx \lor tx)) \end{aligned} \quad (28)$$
- $\exists p_1 \ldots \exists p_n\, F'[p_1, \ldots, p_n] \equiv F'[G_1, \ldots, G_n]$ can be reduced analogously to first-order validity, analogously to P1 for EBUR
- $\exists p\, F'[p, \ldots, p] \models \exists p_1 \ldots \exists p_n\, F'[p_1, \ldots, p_n]$ holds in general

How can $\exists p_1 \ldots \exists p_n\, F'[p_1, \ldots, p_n] \models \exists p\, F'[p, \ldots, p]$ be established?

15. Conclusion

- Preliminary results:
  - Identification of "separability" and "eliminability by uniform replacement" as properties which provide criteria to measure the "elimination power" of elimination methods
  - Suggestion to improve the "elimination power" of DLS-like methods by embedding interpolation-based separation
- Identification of several open issues related to the discussed concepts
- Future issue: taking logics with fixpoint operator into account

16. References

[1] W. Ackermann. "Untersuchungen über das Eliminationsproblem der mathematischen Logik". In: *Math. Ann.* 110 (1935), pp. 390–413.
[2] H. Behmann. "Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem". In: *Math. Ann.* 86.3–4 (1922), pp. 163–229.
[3] J. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, 1983.
[4] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambr. Univ. Press, 2001.
[5] W. Conradie. "On the strength and scope of DLS". In: *JANCL* 16.3–4 (2006), pp. 279–296.
[6] W. Conradie, V. Goranko, and D. Vakarelov. "Algorithmic correspondence and completeness in modal logic. I. The core algorithm SQEMA". In: *LMCS* 2.1:5 (2006), pp. 1–26.
[7] W. Craig. "Linear Reasoning. A New Form of the Herbrand-Gentzen Theorem". In: *J. Symb. Log.* 22.3 (1957), pp. 250–268.
[8] W. Craig. "Three Uses of the Herbrand-Gentzen Theorem in Relating Model Theory and Proof Theory". In: *J. Symb. Log.* 22.3 (1957), pp. 269–285.
[9] W. Craig. "Bases for First-Order Theories and Subtheories". In: *J. Symb. Log.* 25.2 (1960), pp. 97–142.
[10] P. Doherty, W. Łukaszewicz, and A. Szałas. "Computing Circumscription Revisited: A Reduction Algorithm". In: *JAR* 18.3 (1997), pp. 297–338.
[11] D. M. Gabbay, R. A. Schmidt, and A. Szałas. *Second-Order Quantifier Elimination: Foundations, Computational Aspects and Applications*. College Pub., 2008.
[12] D. Gabbay and H. J. Ohlbach. "Quantifier Elimination in Second-Order Predicate Logic". In: *KR'92*. Morgan Kaufmann, 1992, pp. 425–435.
[13] S. Ghilardi, C. Lutz, and F. Wolter. "Did I Damage my Ontology? A Case for Conservative Extensions in Description Logics". In: *KR'06*. AAAI Press, 2006, pp. 187–197.
[14] V. Goranko et al. "SCAN is complete for all Sahlqvist formulae". In: *RelMiCS 7*. Vol. 3051. LNCS. 2004, pp. 149–162.
[15] J. Gustafsson. *An Implementation and Optimization of an Algorithm for Reducing Formulae in Second-Order Logic*. Tech. rep. LiTH-MAT-R-96-04. Univ. Linköping, 1996.
[16] B. Konev, D. Walther, and F. Wolter. "Forgetting and uniform interpolation in large-scale description logic terminologies". In: *IJCAI09*. 2009, pp. 830–835.
[17] R. Kontchakov, F. Wolter, and M. Zakharyaschev. "Logic-based ontology comparison and module extraction, with an application to DL-Lite". In: *AI* 174.15 (2010), pp. 1093–1141.
[18] P. Koopmann and R. A. Schmidt. "Uniform Interpolation of $\mathcal{ALC}$-Ontologies Using Fixpoints". In: *FroCoS 2013*. Vol. 8152. LNCS (LNAI). Springer, 2013, pp. 87–102.
[19] L. Löwenheim. "Über Möglichkeiten im Relativkalkül". In: *Math. Ann.* 76 (4 1915), pp. 447–470.
[20] M. Ludwig and B. Konev. "Practical Uniform Interpolation and Forgetting for $\mathcal{ALC}$ TBoxes with applications to logical difference". In: *KR'14*. AAAI Press, 2014.
[21] C. Lutz and F. Wolter. "Foundations for Uniform Interpolation and Forgetting in Expressive Description Logics". In: *IJCAI-11*. AAAI Press, 2011, pp. 989–995.
[22] R. C. Lyndon. "An interpolation theorem in the predicate calculus". In: *Pacific J. Math.* 9.1 (1959), pp. 129–142.
[23] M. Otto. "An Interpolation Theorem". In: *The Bulletin of Symbolic Logic* 6 (4 200), pp. 447–462.
[24] H. Sahlqvist. "Completeness and correspondence in the first and second order semantics for modal logic". In: *Proc. Third Scand. Logic Symp. Uppsala (1973)*. North-Holland, 1975, pp. 110–143.
[25] R. A. Schmidt. "The Ackermann approach for modal logic, correspondence theory and second-order reduction". In: *JAL* 10.1 (2012), pp. 52–74.
[26] T. Skolem. "Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme welche gewisse Klassen von Aussagen betreffen". In: *Videnskapsselskapets Skrifter* I. Mat.-Nat. Klasse.3 (1919).
[27] A. Szałas. "On the Correspondence between Modal and Classical Logic: An Automated Approach". In: *J. Logic and Comput.* 3 (1993), pp. 605–620.
[28] Z. Wang et al. "Forgetting for knowledge bases in DL-Lite". In: *Ann. Math. Artif. Intell* 58 (2010), pp. 117–151.
[29] C. Wernhard. "Projection and Scope-Determined Circumscription". In: *J. Symb. Comput.* 47.9 (2012), pp. 1089–1108.
[30] C. Wernhard. "Abduction in Logic Programming as Second-Order Quantifier Elimination". In: *FroCoS 2013*. Vol. 8152. LNCS (LNAI). 2013, pp. 103–119.
[31] C. Wernhard. "Second-Order Quantifier Elimination on Relational Monadic Formulas – A Basic Method and Some Less Expected Applications". In: *TABLEAUX 2015*. Vol. 9323. LNCS (LNAI). Springer, 2015.

Acknowledgment: This work was supported by DFG grant WE 5641/1-1