Normalisations of Existential Rules: Not so Innocuous!

David Carral[1], Lucas Larroque[2], Marie-Laure Mugnier[1], Michaël Thomazo[3]

[1]LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France
[2]DI ENS, ENS, CNRS, PSL University, Paris, France
[3]Inria, DI ENS, ENS, CNRS, PSL University, Paris, France

Background

- **Existential rule:** $\forall \vec{x}.\ \forall \vec{y}.\ B[\vec{x}, \vec{y}] \rightarrow \exists \vec{z}.\ H[\vec{x}, \vec{z}]$
- **Datalog rule:** no existential variable ($\vec{z} = \emptyset$)
- **KB:** $\mathcal{K} = (\mathcal{R}, F)$
  with $\mathcal{R}$ and F finite sets of existential rules and of facts

- **Basic query entailment problem**:
  Given a KB $\mathcal{K}$ and a Boolean conjunctive query q,
  does $\mathcal{K} \models q$ hold?
- Two main techniques
  - **Chase** F with $\mathcal{R}$: $\mathcal{K} \models q$ iff $chase(\mathcal{K}) \models q$
  - **Rewrite** q into a FO-query q' such that: $\mathcal{K} \models q$ iff $F \models q'$

Question

Hence, two fundamental properties of rule sets:

- **Chase termination** (for any set of facts F)
- **FO-rewritability** (for any conjunctive query q)

Rule sets are often normalized

Two common procedures:

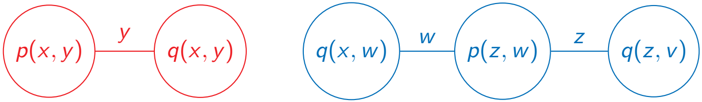
- rule heads decomposed into pieces (**piece decomposition**)
- rule heads decomposed into atoms (**atomic-head decomposition**)

What is their impact on chase termination and FO-rewritability?

Piece of a rule head: maximal subset connected by existential variables

- **Piece graph** of a rule head: one node per atom and an edge AB if A and B share an existential variable
- **Piece**: connected component of the piece graph
- From a rule $R = B \rightarrow H$, we obtain rules $B \rightarrow P_i$, for each piece P_i

$$R = p(x,x) \rightarrow \exists y, z, v, w.\ p(x,y) \wedge q(x,y) \wedge q(x,w) \wedge p(z,w) \wedge q(z,v)$$



We obtain:

- $p(x,x) \rightarrow \exists y.\ p(x,y) \wedge q(x,y)$
- $p(x,x) \rightarrow \exists z, v, w.\ q(x,w) \wedge p(z,w) \wedge q(z,v)$

Normalization: Atomic decomposition

For each rule $R = B[\vec{x}, \vec{y}] \rightarrow \exists \vec{z}.\ H[\vec{x}, \vec{z}]$,
we introduce a fresh predicate X_R
and obtain the rules:

- $B[\vec{x}, \vec{y}] \rightarrow \exists \vec{z}.\ X_R(\vec{x}, \vec{z})$
- $X_R(\vec{x}, \vec{z}) \rightarrow A_i$, for each $A_i \in H$

$R = Manager(x) \rightarrow \exists y.\ \exists z.\ ReportsTo(x, y) \wedge ReportsTo(z, x)$

We obtain:

- $Manager(x) \rightarrow \exists y.\ \exists z.\ X_R(x, y, z)$
- $X_R(x, y, z) \rightarrow ReportsTo(x, y)$
- $X_R(x, y, z) \rightarrow ReportsTo(z, x)$

Back to our question

- Piece decomposition of $\mathcal{R}$: yields a set logically equivalent to $\mathcal{R}$
- Atomic decomposition of $\mathcal{R}$: yields a conservative extension of $\mathcal{R}$

Hence both transformations preserve query entailment

What is their impact on chase termination and FO-rewritability?

FO-rewritability: no impact actually

$\Rightarrow$ *Focus on chase termination*

Chase variants:
Oblivious, Semi-Oblivious, Restricted, Equivalent

Oblivious ($\mathbb{O}$): all rule applications

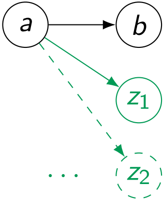
Semi-oblivious ($\mathbb{SO}$): rule applications that differ on the rule frontiers

Similar to **Skolem chase**: all rule applications with skolemized rules

$\mathsf{F} = \{p(a, b)\}$

$p(x, y) \rightarrow \exists z.p(x, z)$

Skolemized rule:
$p(x, y) \rightarrow p(x, f(x))$



$\mathbb{O}$-chase does not terminate, $\mathbb{SO}$-chase does

Chase variants: Restricted

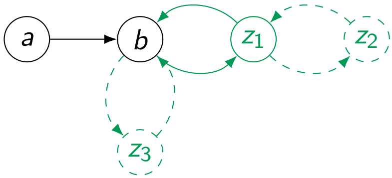
Restricted chase ($\mathbb{R}$): rule applications not already “satisfied”

The added atoms cannot be folded on the previous set of facts

[The homomorphism from the rule body cannot be extended to a homomorphism from the rule head]

$F = \{p(a, b)\}$

$p(x, y) \rightarrow \exists z.p(y, z) \wedge p(z, y)$



$\mathbb{R}$-chase terminates

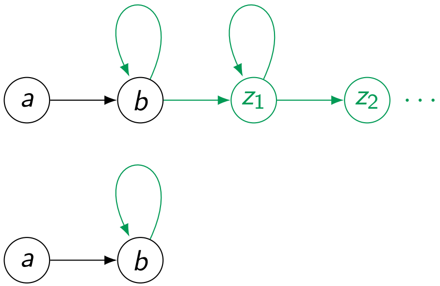
Chase variants: Restricted

For the restricted chase, the order of rule applications matters

$F = \{p(a, b)\}$

$p(x, y) \rightarrow \exists z.p(y, z)$

$p(x, y) \rightarrow p(y, y)$



Datalog-first restricted chase

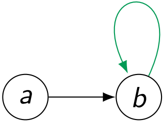
Datalog-first restricted chase ($\mathbb{DF}$-$\mathbb{R}$): priority to Datalog rules

Idea: help folding the atoms of purely existential rules

$F = \{p(a, b)\}$

$p(x, y) \rightarrow \exists z.p(y, z)$

$p(x, y) \rightarrow p(y, y)$



Surprisingly: Datalog-first is not always the best strategy!

Theorem

There are rule sets $\mathcal{R}$ s.t.

- *any KB $(F, \mathcal{R})$ admits a terminating $\mathbb{R}$-chase sequence, but*
- *there is a KB $(F, \mathcal{R})$ without any terminating $\mathbb{DF}$-$\mathbb{R}$-chase sequence.*

Equivalent chase

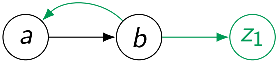
Equivalent chase ($\mathbb{E}$): rule applications s.t. the obtained set of facts is not equivalent to the previous one

Same behavior regarding termination as the **core chase**

$F = \{p(a, b)\}$

$p(x, y) \to \exists z.p(y, z)$

$p(x, y) \wedge p(y, z) \to p(y, x)$



$\mathbb{E}$-chase terminates
while all $\mathbb{R}$-chase sequences are infinite

Classes of rule sets that ensure chase termination

Given a chase variant X

- $\mathcal{R}$ ensures **always termination** if
  for **all** KB $(\mathcal{R}, F)$, **all** (fair) X-chase sequences are finite
  Notation: $\mathcal{R} \in CT^{X}_{\forall\forall}$
- $\mathcal{R}$ ensures **sometimes termination** if
  for **all** KB $(\mathcal{R}, F)$, **at least one** (fair) chase sequence is finite
  Notation: $\mathcal{R} \in CT^{X}_{\forall\exists}$

This distinction is relevant for the restricted chase only

Known inclusions:

$$CT^{\mathbb{O}}_{\forall\forall} \subset CT^{\mathbb{SO}}_{\forall\forall} \subset CT^{\mathbb{R}}_{\forall\forall} \subset CT^{\mathbb{DF}\text{-}\mathbb{R}}_{\forall\forall}$$
$$\subset CT^{\mathbb{DF}\text{-}\mathbb{R}}_{\forall\exists} \subset CT^{\mathbb{R}}_{\forall\exists} \subset CT^{\mathbb{E}}_{\forall\forall}$$

Piece decomposition - Results

	$\mathbb{O}$	$\mathbb{SO}$	$\mathbb{R}$ ($\exists$)	$\mathbb{R}$ ($\forall$)	$\mathbb{DF}$-$\mathbb{R}(\exists)$	$\mathbb{DF}$-$\mathbb{R}(\forall)$	$\mathbb{E}$
Piece decomposition	$=$	$+$	$\neq$	$\neq$	$\neq$	$\neq$	$=$

Chase termination:

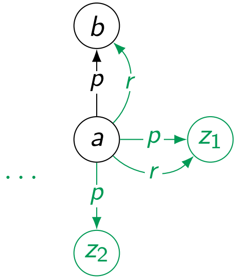
- $+$ can be gained.
- $=$ is unaffected.
- $-$ can be lost.
- $\neq$ can be gained and lost.

Piece decomposition - Results

Restricted chase always-termination can be gained

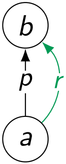
p(a,b)

$p(x, y) \rightarrow \exists z.p(x, z) \wedge r(x, y)$



After piece decomposition:

$p(x, y) \rightarrow \exists z.p(x, z)$
$p(x, y) \rightarrow r(x, y)$

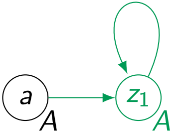


Piece decomposition - Results

Restricted chase always-termination can be lost

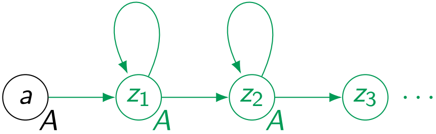
A(a)

$A(x) \rightarrow \exists z.p(x,z)$
$p(x,y) \rightarrow p(y,y) \wedge A(y)$



After piece decomposition:

$A(x) \rightarrow \exists z.p(x,z)$
$p(x,y) \rightarrow p(y,y)$
$p(x,y) \rightarrow A(y)$



Atomic decomposition - Results

	$\mathbb{O}$	$\mathbb{SO}$	$\mathbb{R}$ $(\exists)$	$\mathbb{R}$ $(\forall)$	$\mathbb{DF}$-$\mathbb{R}(\exists)$	$\mathbb{DF}$-$\mathbb{R}(\forall)$	$\mathbb{E}$
One-way atomic decomposition	$=$	$=$	$-$	$-$	$-$	$-$	$-$

Impact on termination:

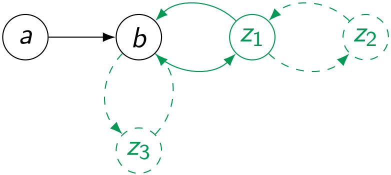
- $+$ can be gained.
- $=$ is unaffected.
- $-$ can be lost.
- $\neq$ can be gained and lost.

Atomic decomposition - Results

Restricted chase always-termination can be lost

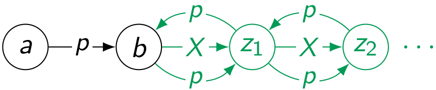
$p(a, b)$

$p(x, y) \rightarrow \exists z.p(y, z) \wedge p(z, y)$



After normalization:

$p(x, y) \rightarrow \exists z.X(y, z)$
$X(y, z) \rightarrow p(y, z)$
$X(y, z) \rightarrow p(z, y)$



Two-way atomic decomposition

For each rule $R = B[\vec{x}, \vec{y}] \to \exists \vec{z}.\ H[\vec{x}, \vec{z}]$,
we introduce a fresh predicate X_R
and obtain the rules:

- $B[\vec{x}, \vec{y}] \to \exists \vec{z}.\ X_R(\vec{x}, \vec{z})$
- $X_R(\vec{x}, \vec{z}) \to A_i$, for each $A_i \in H$
- $H \to X_R(\vec{x}, \vec{z})$

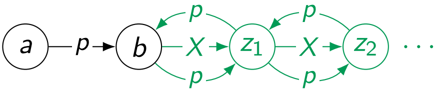
$R = Manager(x) \to \exists y.\ \exists z.\ ReportsTo(x, y) \land ReportsTo(z, x)$
yields:

- $Manager(x) \to \exists y.\ \exists z.\ X_R(x, y, z)$
- $X_R(x, y, z) \to ReportsTo(x, y)$
- $X_R(x, y, z) \to ReportsTo(z, x)$
- $ReportsTo(x, y) \land ReportsTo(z, x) \to X_R(x, y, z)$

Two-way helps $\mathbb{R}$-chase sometimes-termination

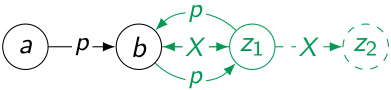
After one-way decomposition:

$p(x, y) \rightarrow \exists z.X(y, z)$
$X(y, z) \rightarrow p(y, z)$
$X(y, z) \rightarrow p(z, y)$



After two-way decomposition:

$p(x, y) \rightarrow \exists z.X(y, z)$
$X(y, z) \rightarrow p(y, z)$
$X(y, z) \rightarrow p(z, y)$
$p(y, z) \wedge p(z, y) \rightarrow X(y, z)$



Two-way atomic decomposition - Results

	$\mathbb{O}$	$\mathbb{SO}$	$\mathbb{R}$ $(\exists)$	$\mathbb{R}$ $(\forall)$	$\mathbb{DF}$-$\mathbb{R}(\exists)$	$\mathbb{DF}$-$\mathbb{R}(\forall)$	$\mathbb{E}$
One-way atomic decomposition	=	=	−	−	−	−	−
Two-way atomic decomposition	=	=	+	−	=	=	=

Two-way improves over one-way:

- No impact on $\mathbb{DF}$-$\mathbb{R}$-chase nor $\mathbb{E}$-chase
- Positive impact on sometimes-termination of $\mathbb{R}$-chase
- ... but still a negative impact on always-termination of $\mathbb{R}$-chase

No better atomic decomposition?

There is no **computable** atomic decomposition that **exactly preserves** $\mathbb{R}$-chase always-termination.

$CT^{\mathbb{R}}_{F\forall}$, with F fixed: class of rule sets $\mathcal{R}$ s.t. all $\mathbb{R}$-chase sequences from the KB $(\mathcal{R}, F)$ are finite

Theorem

- *For all F, the subset of $CT^{\mathbb{R}}_{F\forall}$ restricted to atomic-head rules is recursively enumerable*
- *There is F such that $CT^{\mathbb{R}}_{F\forall}$ is not recursively enumerable (Π^2_0-hard)*

Hence, no computable function f exists that maps rule sets to atomic-head rule sets s.t., for all F and $\mathcal{R}$, $\mathcal{R} \in CT^{\mathbb{R}}_{F\forall}$ iff $f(\mathcal{R}) \in CT^{\mathbb{R}}_{F\forall}$

Conclusions

	$\mathbb{O}$	$\mathbb{SO}$	$\mathbb{R}$ ($\exists$)	$\mathbb{R}$ ($\forall$)	$\mathbb{DF}$-$\mathbb{R}$ ($\exists$)	$\mathbb{DF}$-$\mathbb{R}$ ($\forall$)	$\mathbb{E}$	FO Rewritability
Piece	$=$	$+$	$\neq$	$\neq$	$\neq$	$\neq$	$=$	$=$
1-way atomic	$=$	$=$	$-$	$-$	$-$	$-$	$-$	$=$
2-way atomic	$=$	$=$	$+$	$-$	$=$	$=$	$=$	$=$

Further results on the restricted chase:

- Datalog-first strategy is not optimal regarding termination
- Termination of $\mathbb{R}$-chase on a KB is not recursively enumerable (while it is for atomic-head rules)

Future Work

1. Develop normalisation procedures that (A) transform FO-theories into sets of disjunctive existential rules and (B) preserve query entailment (over the original signature), chase termination, and FO-rewritability

2. The theorem that we proved:

Theorem

*There is no **computable** atomic decomposition that **exactly preserves** $\mathbb{R}$-chase always-termination.*

The one that we wanted is obtained by removing "exactly" above. This result would follow from:

Hypothesis. The subset of $CT^{\mathbb{R}}_{\forall\forall}$ restricted to atomic-head rules is recursively enumerable

Thank you for your attention!