A Systematic Approach to Deriving Incremental Type Checkers

ANDRÉ PACAK, JGU Mainz, Germany
SEBASTIAN ERDWEG, JGU Mainz, Germany
TAMÁS SZABÓ, JGU Mainz / itemis, Germany

Static typing can guide programmers if feedback is immediate. Therefore, all major IDEs incrementalize type checking in some way. However, prior approaches to incremental type checking are often specialized and hard to transfer to new type systems. In this paper, we propose a systematic approach for deriving incremental type checkers from textbook-style type system specifications. Our approach is based on compiling inference rules to Datalog, a carefully limited logic programming language for which incremental solvers exist. The key contribution of this paper is to discover an encoding of the infinite typing relation as a finite Datalog relation in a way that yields efficient incremental updates. We implemented the compiler as part of a type system DSL and show that it supports simple types, some local type inference, operator overloading, universal types, and iso-recursive types.

CCS Concepts: • **Theory of computation** → **Type structures**; **Program analysis**; *Constraint and logic programming*.

Additional Key Words and Phrases: incremental type checking, datalog, type system transformation

ACM Reference Format:
André Pacak, Sebastian Erdweg, and Tamás Szabó. 2020. A Systematic Approach to Deriving Incremental Type Checkers. *Proc. ACM Program. Lang.* 4, OOPSLA, Article 127 (November 2020), 28 pages. https://doi.org/10.1145/3428195

1 INTRODUCTION

Many programming languages employ a static type system to check user-defined invariants at compile time. Indeed, programmers of statically typed languages often rely on feedback from the type checker for guidance. Unfortunately, type checking can take significant time for larger programs and can interrupt the programmer's development flow. Therefore, it is hardly surprising that virtually all major IDEs incrementalize type checking in some way. Unfortunately, most of these solutions are highly specialized and generally hard to transfer to a new type system. We lack a principled solution for incrementalizing type checkers.

This paper presents a systematic approach for *deriving* incremental type checkers from textbook-style type system specifications. While we hope to incrementalize advanced type systems eventually, in this paper we focus on building a sound foundation for the most elemental type system features: name binding and type errors. These features constitute fundamental challenges for incremental type checking, and we believe it is essential to develop principled solutions to such fundamental challenges first. Nonetheless, our approach already supports a number of sophisticated type systems.

Authors' addresses: André Pacak, JGU Mainz, Germany; Sebastian Erdweg, JGU Mainz, Germany; Tamás Szabó, JGU Mainz / itemis, Germany.

2475-1421/2020/11-ART127
https://doi.org/10.1145/3428195

Our approach is based on the idea of compiling inference rules to the logic programming language Datalog. Targeting Datalog is promising because efficient incremental Datalog solvers already exist [Ujhelyi et al. 2015]. However, targeting Datalog is also challenging because Datalog's expressivity is carefully limited. Datalog programs can only compute finite relations, whereas the typing relation usually is an inductively defined infinite relation. Although this makes compiling type checkers to Datalog seemingly impossible, we have discovered a sequence of systematic transformations that make the resulting inference rules expressible in Datalog.

The first transformation utilizes a new property we call *co-functional dependencies*. While a functional dependency describes a uniquely determined output, a co-functional dependency describes a uniquely determined input. In particular, for algorithmic type systems, the typing context and other contextual information is co-functionally dependent on the syntax tree. Our transformation exploits this property to factor out the context from the typing relation, making the typing relation computable in Datalog. Unfortunately, the resulting type system won't admit efficient incrementalization, because even a small change to the typing context will affect large parts of the typing derivation. We discovered that we can eliminate this issue by complete deforestation [Wadler 1990] of all typing contexts. Thus, our second transformation is a specialized deforestation of Datalog programs. Our third and final transformation makes sure ill-typed terms do not unnecessarily prune typing derivations. Otherwise, any code change that fixes a type error would entail significant reanalysis. To this end, we developed a reformulation of type systems that separates error handling from computing a type. Our transformation rewrites any algorithmic type checker into one that collects type errors separately from the typing relation. This transformation may well be useful independent of the rest of our work.

Based on our transformations, we developed a domain-specific language (DSL) for type system descriptions that compiles to Datalog. Our compiler implements the three transformations by rewriting inference rules and only lowers those inference rules to Datalog as the last step. Thus, the DSL compiler can be used to simplify type systems independent of Datalog. We have used the DSL to express a wide range of type system features. In addition to PCF with product and sum types, we modeled bi-directional type checking, operator overloading, universal types in the style of System F, and iso-recursive types. We can confirm that all these features can be compiled to Datalog by our transformations and that the resulting Datalog program is incrementally solvable. Effectively, our DSL derives incremental type checkers from textbook-like type system specifications. We also measured the incremental performance of compiled type systems for synthesized PCF programs. We designed a range of change scenarios to challenge the incremental performance. We found that even when large parts of the program are affected by a change, we still deliver updated typing information in at most several tens of milliseconds.

In summary, we make the following contributions:

- We analyze the challenges associated with compiling type systems to Datalog (Section 2).
- We introduce co-functional dependencies and define a Datalog transformation that moves co-functionally dependent data into a separate relation (Section 3).
- We show how to eliminate typing context propagation from type systems (Section 4).
- We show how to transform a type system to collect type errors on the side (Section 5).
- We implement all three transformations in the compiler of a type systems DSL (Section 6), demonstrate its applicability (Section 7), and benchmark its performance (Section 8).

2 WHY ARE TYPE SYSTEMS IN DATALOG CHALLENGING?

This paper proposes to incrementalize type checkers by translation to Datalog. Our hypothesis is that such translation can be done systematically and is useful: Existing incremental Datalog solvers provide efficient incremental running times. In the present section, we illustrate why encoding type checkers in Datalog is challenging in the first place. We highlight the challenges while translating a number of exemplary type systems, all using the following syntax:

$$
\begin{array}{lll}
\text{(program)} & p & ::= \text{main } e \\
\text{(expression)} & e & ::= \text{unit} \mid x \mid \lambda x{:}T.\, e \mid e\ e \\
\text{(type)} & T & ::= \text{Unit} \mid T \to T \\
\text{(context)} & C & ::= \varepsilon \mid C, x{:}T
\end{array}
$$

Our motivating examples will only differ in their typing relation but reuse the same syntax. For each typing relation, we show how the type rules can be translated to Datalog and discuss if and how a state-of-the-art incremental Datalog solver could handle them. As such, the current section also presents the required Datalog background.

Challenge 1: Expressions. We start with the typing relation $(e : T)$ of a very simple type system that only permits unit constants and their application. This type system is not particularly useful but helps us illustrate how to translate a simple type system to Datalog.

$$
\text{T-Unit}\,\frac{}{\text{unit} : \text{Unit}} \qquad \text{T-App}\,\frac{e_1 : \text{Unit} \quad e_2 : \text{Unit}}{e_1\ e_2 : \text{Unit}} \qquad \text{T-Main}\,\frac{e : T}{(\text{main } e)\ \text{ok}}
$$

We can represent the typing relation $(e : T)$ as a binary Datalog relation $\texttt{typed}(e, T)$ and translate each inference rule to a Datalog rule as follows:

$$
\begin{array}{rl}
\texttt{typed}(e, T) & \text{:- } \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T). \\
\texttt{typed}(e, T) & \text{:- } \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(e_1, T_1),\ \ ?\texttt{Unit}(T_1),\ \ \texttt{typed}(e_2, T_2),\ \ ?\texttt{Unit}(T_2),\ \ !\texttt{Unit}(T). \\
\texttt{ok}(p) & \text{:- } \ ?\texttt{main}(p, e),\ \ \texttt{typed}(e, T).
\end{array}
$$

A Datalog program consists of a sequence of rules, each of the form $\mathtt{R}(t_1, \ldots, t_n)$:- $\mathtt{a}_1, \ldots, \mathtt{a}_\mathtt{m}$. The rule head $\mathtt{R}(t_1, \ldots, t_n)$ declares tuple $(t_1, \ldots, t_n) \in \mathtt{R}$ if all atoms $\mathtt{a}_1, \ldots, \mathtt{a}_\mathtt{m}$ in the rule body hold. Terms t are usually logical variables that are shared between the head and body of a rule. Atoms a query relations $\mathtt{R}(t_1, \ldots, t_n)$. This way, relations can depend on each other recursively. In this paper, we use Datalog enriched with algebraic data types in order to model expressions, types, contexts, etc. For each constructor $c(x_1, \ldots, x_k)$ of an algebraic data type, we assume operations $!\mathtt{c}(y, x_1, \ldots, x_k)$ and $?\mathtt{c}(y, x_1, \ldots, x_k)$ to construct and deconstruct algebraic data x.

Given this Datalog background, it should be easy to see that $(e, T) \in \texttt{typed}$ if and only if there is a derivation tree for $(e : T)$ according to the inference rules. We hope to incrementalize type checking (i.e., finding a derivation tree) by applying existing incremental Datalog solvers to the derived Datalog program. To this effect, it is important to know that incremental Datalog solvers evaluate Datalog rules *bottom-up*, inductively enumerating *all derivable tuples*. When the input changes, an incremental Datalog solver updates the relations by retracting those tuples no longer derivable and inserting the newly derivable tuples. Unfortunately, this strategy hinges on the Datalog relations being finite. However, our typing relation is infinite, because our example language contains infinitely many well-typed programs:

$$
\texttt{typed} = \{(\text{unit}, \text{Unit}), ((\text{unit unit}), \text{Unit}), (((\text{unit unit})\ \text{unit}), \text{Unit}), \ldots\}
$$

Dealing with an infinite language is a standard problem when using Datalog for static analysis. Even Datalog-based analysis systems without incrementalization such as Doop [Smaragdakis and Bravenboer 2010] require finite relations. Fortunately, there is a standard solution that we can employ here as well: Restrict the relations to only consider the user's current program. That is, rather than defining $?\texttt{unit}$, $?\texttt{app}$, and $?\texttt{main}$ inductively over all possible programs, we define them as sets of tuples that exactly reflect the user program. Since the user program is finite by construction, we can now evaluate `typed` bottom-up, enumerating the well-typed subset of the nodes in ?`unit`, ?`app`, and ?`main`. This strategy has been successfully employed in Datalog-based incremental analyzers before [Szabó et al. 2016], such that no further innovation is required for this first challenge. As an example we show how the program on the left is encoded in Datalog as you can see on the right:

$$\overbrace{\text{main}\ (\overbrace{\underbrace{(\underbrace{\text{unit}}_{e3}\ \ \underbrace{\text{unit}}_{e4})}_{e2}\ \ \underbrace{\text{unit}}_{e5})}^{e1}}^{p}
\qquad\qquad
\begin{aligned}
?\texttt{main} &= \{(p, e1)\}\\
?\texttt{app} &= \{(e1, e2, e5), (e2, e3, e4)\}\\
?\texttt{unit} &= \{e3, e4, e5\}
\end{aligned}$$

We use labels such as p and $e1$ in the Datalog encoding to reference nodes of the abstract syntax tree. The set of tuples ?`app` contains tuples that represent application nodes. Take for example the tuple $(e1, e2, e5)$ where node $e1$ is an application node which has $e2$ and $e5$ as children. A Datalog solver will derive the following finite relations based on the Datalog rules above and the sets of tuples:

$$\begin{aligned}
\texttt{typed} &= \{(e3, \text{Unit}), (e4, \text{Unit}), (e5, \text{Unit}), (e2, \text{Unit}), (e1, \text{Unit})\}\\
\texttt{ok} &= \{p\}
\end{aligned}$$

Challenge 2: Types. The previous type system is not very useful because it only inhabits the Unit type. We extend this type system by allowing thunks and their application:

$$\text{T-Unit}\,\frac{}{\text{unit} : \text{Unit}} \qquad \text{T-App}\,\frac{e_1 : \text{Unit} \rightarrow T \quad e_2 : \text{Unit}}{e_1\ e_2 : T}$$

$$\text{T-Lam}\,\frac{e_1 : T_2}{\lambda x{:}\text{Unit}.\ e_1 : \text{Unit} \rightarrow T_2} \qquad \text{T-Main}\,\frac{e : T}{(\text{main}\ e)\ \text{ok}}$$

Again, we can translate these type rules to Datalog as explained above:

$$\begin{aligned}
\texttt{typed}(e, T) &\ \texttt{:-}\ ?\texttt{unit}(e),\ !\texttt{Unit}(T).\\
\texttt{typed}(e, T) &\ \texttt{:-}\ ?\texttt{app}(e, e_1, e_2),\ \texttt{typed}(e_1, T_e),\ ?\texttt{Fun}(T_e, T_1, T),\ ?\texttt{Unit}(T_1),\ \texttt{typed}(e_2, T_2),\\
&\qquad ?\texttt{Unit}(T_2).\\
\texttt{typed}(e, T) &\ \texttt{:-}\ ?\texttt{lam}(e, x, T_1, e_1),\ ?\texttt{Unit}(T_1),\ \texttt{typed}(e_1, T_2),\ !\texttt{Fun}(T, T_1, T_2).\\
\texttt{ok}(p) &\ \texttt{:-}\ ?\texttt{main}(p, e),\ \texttt{typed}(e, T).
\end{aligned}$$

Again, we must ask if these rules can be evaluated bottom-up by an incremental Datalog solver. And again this hinges on the Datalog relations being finite. If we assume like above that ?`app`, ?`lam`, etc. are constants and only enumerate the user's current program, then only finitely many expressions can occur in `typed`. Indeed, expressions are not the problem but types are. While the previous type system only considered a single type Unit, this type system associates thunk types of the form $T ::= \text{Unit} \mid \text{Unit} \rightarrow T$ to expressions. Notably, this domain is infinite and relation $\texttt{typed} \subseteq \texttt{e}_{user} \times T$ could contain infinitely many tuples even if $\texttt{e}_{user}$ is finite.

In truth, `typed` will only ever contain finitely many tuples. This is because of the functional dependency $e \rightsquigarrow T$ in `typed`, which means that column T of `typed` is uniquely determined by column e of `typed`. That is, if $(e, T1) \in \texttt{typed}$ and $(e, T2) \in \texttt{typed}$, then $T1 = T2$ [Watt 2018, Chap. 11]. Our type system satisfies the functional dependency $e \rightsquigarrow T$ because it is algorithmic. Consequently, if `typed` only contains finitely many entries in column e, then `typed` can also only contain finitely many tuples in $e \times T$. Therefore, `typed` is finite and incremental bottom-up evaluation succeeds [Ramakrishnan et al. 1987].

It is noteworthy that many (even non-incremental) Datalog solvers would reject the derived Datalog program because it synthesizes data at run time. However, our type system must generate function types $!\texttt{Fun}(T, T_1, T_2)$ of arbitrary size to match the nesting level of lambdas in the user's program. The good news is that a few cutting-edge incremental Datalog solvers like IncA [Szabó et al. 2018a] and DDlog [Ryzhyk and Budiu 2019] can handle the derived Datalog code. The bad news is that we must move beyond the cutting edge to support more interesting type systems.

Challenge 3: Contexts. The next challenge arises when introducing typing contexts. To this end, we consider the simply typed lambda calculus:

$$\text{T-Unit}\,\frac{}{C \vdash \text{unit} : \text{Unit}} \qquad \text{T-App}\,\frac{C \vdash e_1 : T_1 \to T \quad C \vdash e_2 : T_1}{C \vdash e_1\ e_2 : T}$$

$$\text{T-Lam}\,\frac{C, x{:}T_1 \vdash b : T_2}{C \vdash \lambda x{:}T_1.\, b : T_1 \to T_2} \qquad \text{T-Var}\,\frac{C(x) = T}{C \vdash x : T} \qquad \text{T-Main}\,\frac{\varepsilon \vdash e : T}{(\text{main } e)\ \text{ok}}$$

The typing relation now is ternary and the inference rules thread the typing context:

$$\begin{array}{rl}
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T).\\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(C, e_1, T_e),\ \ ?\texttt{Fun}(T_e, T_1, T),\ \ \texttt{typed}(C, e_2, T_1).\\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{lam}(e, x, T_1, b),\ \ !\texttt{bind}(C', C, x, T_1),\ \ \texttt{typed}(C', b, T_2),\ \ !\texttt{Fun}(T, T_1, T_2).\\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{var}(e, x),\ \ \texttt{lookup}(C, x, T).\\
\texttt{ok}(p) & \texttt{:-}\ \ ?\texttt{main}(p, e),\ \ !\texttt{empty}(C),\ \ \texttt{typed}(C, e, T).
\end{array}$$

Note that we use !`bind` in the `lam` case to extend the context and `lookup` in the `var` case to extract a binding from the context. The main program is checked in the !`empty` context.

Unfortunately, our derived Datalog program is not computable in bottom-up style anymore and, thus, cannot be incrementalized by existing Datalog solvers. To see why, let us inspect the `var` rule in more detail. This rule declares a tuple $(C, e, T) \in \texttt{typed}$ whenever $?\texttt{var}(e, x)$ and $\texttt{lookup}(C, x, T)$ hold. As argued above, we can restrict e to range over the user's program only, so that only finitely many variable symbols have to be considered here. But unlike before, e does no longer uniquely determine T because the type also depends on the context C. Therefore, even a single variable x has infinitely many potential typing derivations:

$$\begin{aligned}\texttt{typed} = \{\ & (x{:}\text{Unit}, x, \text{Unit}),\quad (x{:}\text{Unit} \to \text{Unit}, x, \text{Unit} \to \text{Unit}),\\ & (x{:}(\text{Unit} \to \text{Unit}) \to \text{Unit}, x, (\text{Unit} \to \text{Unit}) \to \text{Unit}),\ \ldots\}\end{aligned}$$

As we will show in Section 3, a different encoding of type systems in Datalog can solve this problem. Our solution works for algorithmic type systems and is based on the following observations:

(1) Algorithmic type systems do not guess substitutions of metavariables, but require metavariables to be positively bound. In particular, when a judgment $\texttt{typed}(C, e, T)$ occurs as a premise, the context C is uniquely determined.
(2) Algorithmic type systems are syntax-directed and conduct a fold over the syntax tree. This means that each node in the syntax tree is visited at most once during typing.

Together, these observations entail that each expression is checked under a single, uniquely determined context. We exploit this to factor out the context from relation `typed`, adding a new relation $\texttt{context}(e, C)$ that associates contexts to expressions. Both relations are finite now. In particular, each variable x in the syntax tree occurs in a unique context $\texttt{context}(x, C)$ and therefore has a unique type $\texttt{typed}(x, T)$.

Challenge 4: Context propagation. By factoring out the context from relation `typed`, we obtained a Datalog program that is computable in bottom-up style. Thus, we can apply cutting-edge incremental Datalog solvers like IncA [Szabó et al. 2018a] and DDlog [Ryzhyk and Budiu 2019] to it. Unfortunately, this will yield unsatisfactory incremental performance. In general, an incremental algorithm yields good incremental performance if the *size of a change* correlates with the *time it takes to process that change*. Conversely, the update time should be largely independent of the size of the overall input. However, for our derived Datalog code, many small changes in the user program can require a large amount of reanalysis. This problem is due to context propagation.

Consider the following example program, where we use let as syntactic sugar:

$$\begin{array}{l}\text{let } id : \text{Unit} \rightarrow \text{Unit} = \lambda x{:}\text{Unit}.\, x \\ \text{ in } \lambda y{:}\text{Unit}.\, \lambda z{:}\text{Unit}.\, e_0\end{array}$$

Expression e_0 will be checked in a typing context that binds id, y, and z. Now, if the type of id changes in any way, all previously propagated contexts have to be retracted and new contexts have to be propagated. Specifically, the tuples of relation $\texttt{context}(e, C)$ become obsolete for all expressions e where id is in scope, even for expressions that do not actually refer to id. For declarations with a wide scope, such as top-level functions, this behavior will incur a significant incremental performance penalty.

The problem is that the derived Datalog code propagates entire contexts rather than individual context bindings, and that it ignores whether a binding is being used. As we will show in Section 4, we can systematically transform the Datalog code to solve this problem. To do so, we will generate a new relation $\texttt{findBinding}(x, e, T)$ that finds the bound type of variable x occurring within expression e. This relation will walk the syntax tree in the opposite direction of context propagation until a binder for x is found. Since `findBinding` does not require a context, we will be able to drop relation `context` and rewrite the `var` rule as follows:

$$\texttt{typed}(e, T) \ \text{:-}\ \ ?\texttt{var}(e, x),\ \ \texttt{findBinding}(x, e, T).$$

That is, starting at the reference e, we find the bound type of x. With this change, no context propagation will be necessary anymore.

Challenge 5: Ill-typed terms. Static type systems restrict the syntactically well-formed terms and define a language of well-typed programs. A syntactically well-formed term is well-typed if there is a typing derivation for that term. This is a yes or no decision: in or out. For an algorithmic type system, as soon as any rule fails to satisfy a premise, the entire program is known to be ill-typed and typing can stop right there. However, aborting type checking early is unsatisfactory for Datalog-based incrementality and for programmer feedback.

For Datalog-based incrementality, aborting type checking is unsatisfactory since it prunes tuples from the typing relation unnecessarily. In particular, any typing that transitively depends on an ill-typed term will be dropped from the typing relation `typed`. For a simple example, consider a term using type ascription (e as T). If e is ill-typed, e and all its ancestors will be dropped from `typed` because the type rules require subterms to be well-typed. However, notice how the type of (e as T) really is independent of the well-typedness of e. We would like to retain $(e, T) \in \texttt{typed}$, which also allows the ancestors to be checked as usual. As a developer makes changes in quick succession, alternating between a well-typed and an ill-typed program, a more stable `typed` relation means faster update times.

The second concern with aborting type checking at the first type error is that this is inconvenient in practice. Both compilers and programming editors usually try to report all type errors in the program. In Section 5, we show that the Datalog code can be systematically rewritten to collect all type errors and to avoid pruning the typing relation. To this end, we will generate another relation `errors`(e, err) that associates type errors to expressions. A program p then is only well-typed if $p \in \texttt{ok}$ and $\texttt{errors} = \emptyset$.

Problem Statement. The goal of this paper is to translate algorithmic type systems to Datalog to utilize state-of-the-art incremental Datalog solvers. The translation should be systematic, applicable to a wide range of type systems, and yield good incremental performance. In this section, we identified the following five challenges:

(C1) Expressions are drawn from an infinite domain.
(C2) Types are drawn from an infinite domain.
(C3) Contexts are drawn from an infinite domain.
(C4) Contexts are threaded through typing derivations.
(C5) Ill-typed subterms abort type checking.

While prior work on Datalog-based static analysis can be used to solve challenges **(C1)** and **(C2)**, the other challenges require novel solutions. We present Datalog transformations that solve challenges **(C3)**–**(C5)** in Sections 3–5.

Algorithmic Type Systems. In this paper, we assume type systems are given in algorithmic form. Specifically, challenges **(C2)** and **(C3)** require an algorithmic formulation. The word *algorithmic* means that the type system can be trivially translated into a recursive-descent algorithm, but the type system may still be defined using inference-style type rules. Most standard type checkers are naturally algorithmic: Type rules are syntax-directed and metavariables are positively bound. We therefore argue that requiring an algorithmic type system is a modest restriction.

To support our point, note that most textbooks introduce type systems in an algorithmic style, including Harper's *Practical Foundations for Programming Languages* [Harper 2016] and Pierce's *Types and Programming Languages* [Pierce 2002]. Hence, novice language designers learn about type systems by being introduced to their algorithmic formulation. Only complex type system features are sometimes given non-algorithmically and require a rewriting. However, the same rewriting is necessary for non-incremental implementations of those type checkers [Grewe et al. 2015], meaning we do not actually impose an additional requirement on type system designers.

3 TRANSFORMATION 1: CO-FUNCTIONAL DEPENDENCIES

Incremental Datalog solvers evaluate Datalog programs bottom-up. In the previous section, we explained why a naive translation of a type system to Datalog does not permit the application of bottom-up Datalog solvers (Challenge 3): Since contexts occur as a column in the typing relation `typed`, the typing relation has infinitely many tuples, as we illustrated for the `var` rule. Our solution to Challenge 3 is based on a property of algorithmic type systems that we discovered and named *co-functional dependencies.*

3.1 Co-Functional Dependencies

Co-functional dependencies express uniqueness relationships between columns of a relation, similar to functional dependencies. Intuitively, a functional dependency describes unique "outputs" of a relation, whereas a co-functional dependency describes unique "inputs" of a relation. For example, the typing relation of the simply typed lambda calculus ($\texttt{typed} \subseteq C \times e \times T$) has a functional dependency $(C \times e) \rightsquigarrow T$. That is, given C and e, type T is an "output" of typing that is uniquely determined by $C \times e$. Our new observation is that the typing relation also has a co-functional dependency $e \overset{co}{\rightsquigarrow} C$. That is, given e, context C is an "input" of typing that is uniquely determined by e and how `typed` is used. While the treatment of functional dependencies is standard in databases [Watt 2018, Chap. 11] and Datalog [Ramakrishnan et al. 1987], our notion of co-functional dependencies is novel to the best of our knowledge. Unfortunately, co-functional dependencies are harder to detect and utilize, since they depend on how a relation is being used.

The typing context C of the simply typed lambda calculus is an example of a co-functionally dependent column: Each expression is only checked under a single context. We do not know how to detect co-functional dependencies automatically, but instead rely on domain knowledge about algorithmic type systems. In general, all contextual information passed around in an algorithmic type system is uniquely determined for a syntax-tree node. This is because each syntax-tree node is visited at most once per relation (syntax-directedness), and the relevant context information is unique (no guessing of metavariables). We could in principle also allow multiple visits of the same syntax-tree node as long as relevant context information is identical in all visits. For example, this will allow us to support operator overloading with overlapping inference rules (Section 7).

In the remainder of this paper, we assume functional and co-functional dependencies are declared as part of a relation's signature. To this end, we introduce the following notation for signatures:

$$\begin{array}{lll}
\text{(relation signature)} & \sigma & ::= R : T_1 \times \ldots \times T_n \,|\, F, G \\
\text{(functional dependencies)} & F & ::= \{\mathcal{P}(\mathbb{N}) \rightsquigarrow \mathbb{N}, \ldots\} \\
\text{(co-functional dependencies)} & G & ::= \{\mathcal{P}(\mathbb{N}) \overset{co}{\rightsquigarrow} \mathbb{N}, \ldots\}
\end{array}$$

A relation signature $(R : T_1 \times \ldots \times T_n \,|\, F, G)$ describes the columns of relation R, its functional dependencies F, and its co-functional dependencies G. Functional and co-functional dependencies are defined based on column indices. For example, we can represent the typing relation of the simply typed lambda calculus by signature $(\texttt{typed} : C \times e \times T \,|\, \{\{1, 2\} \rightsquigarrow 3\}, \{\{2\} \overset{co}{\rightsquigarrow} 1\})$. The functional dependency declares that columns 1 and 2 together uniquely determine column 3, that is, $C \times e \rightsquigarrow T$. The co-functional dependency declares that column 2 also uniquely determines column 1, that is, $e \overset{co}{\rightsquigarrow} C$. Datalog relations annotated this way enable us to utilize co-functional dependencies.

Notation. We frequently need to denote sequences and subsequences in this paper. We write $\overline{x}$ or $x_1, \ldots, x_n$ for a sequence of x elements. Given a set of indices I, we write x_I for the subsequence of $\overline{x}$ consisting of $\{x_i \mid i \in I\}$ and ordered by their index. We leniently write $\overline{x}, y$ and x_I, y and x_I, x_J to concatenate sequences and sequence elements.

3.2 Utilizing Co-Functional Dependencies

A co-functional dependency $\overline{c} \overset{co}{\rightsquigarrow} c$ in relation R stipulates that column c of R is uniquely determined by some other columns $\overline{c}$ of R. This allows us to factor out c from R, since we can always use the other columns $\overline{c}$ to uniquely obtain c. However, the rules to obtain c from $\overline{c}$ are not obvious and depend on how R is being queried. This makes co-functional dependencies difficult to utilize.

We have developed a transformation of Datalog code that factors out co-functionally dependent columns c from their relation R. The key idea is to derive an auxiliary relation $\pi_R : \overline{c} \times c$ that has a (regular) functional dependency $\overline{c} \rightsquigarrow c$. Essentially, π_R witnesses the contextual uniqueness of c by mapping $\overline{c}$ to c locally. We then rewrite R to drop column c and to query π_R instead. Essentially, if $(\mathtt{R}(\overline{x}, \overline{c}, c) \text{ :- } a)$ is a rule of R, then $(\mathtt{R}'(\overline{x}, \overline{c}) \text{ :- } \pi_{\mathtt{R}}(\overline{c}, c), a)$ will be a rule of the rewritten R'.

Before delving into the technical details of the transformation, let us consider its application to the simply typed lambda calculus whose Datalog rules we showed in Section 2. Since $\texttt{typed} : C \times e \times T$ has $e \overset{co}{\rightsquigarrow} C$, we derive the auxiliary relation $\pi_{\texttt{typed}} : e \times C$ and use it in $\texttt{typed}$:

$$\begin{array}{rl}
\texttt{typed}(e, T) & \text{:- } \pi_{\texttt{typed}}(e, C),\ \texttt{?unit}(e),\ \texttt{!Unit}(T). \\
\texttt{typed}(e, T) & \text{:- } \pi_{\texttt{typed}}(e, C),\ \texttt{?app}(e, e_1, e_2),\ \texttt{typed}(e_1, T_e),\ \texttt{?Fun}(T_e, T_1, T),\ \texttt{typed}(e_2, T_1). \\
\texttt{typed}(e, T) & \text{:- } \pi_{\texttt{typed}}(e, C),\ \texttt{?lam}(e, x, T_1, b),\ \texttt{!bind}(C', C, x, T_1),\ \texttt{typed}(b, T_2),\ \texttt{!Fun}(T, T_1, T_2). \\
\texttt{typed}(e, T) & \text{:- } \pi_{\texttt{typed}}(e, C),\ \texttt{?var}(e, x),\ \texttt{lookup}(C, x, T). \\
\texttt{ok}(p) & \text{:- } \texttt{?main}(p, e),\ \texttt{!empty}(C),\ \texttt{typed}(e, T).
\end{array}$$

Note how we dropped column C from all rule heads and usages of `typed`. Instead, we introduced the query $\pi_{\texttt{typed}}(e, C)$ at the beginning of each `typed` rule to bind C.

For the derived relation $\pi_{\texttt{typed}} : e \times C$, we generate one rule for each call of `typed`. Thus, there is no rule for `unit` because its rule does not call `typed`, but there are two rules for `app`. The derived rules reflect how the co-functionally dependent input C was constrained. Essentially, for each call of $\texttt{typed}(C, e, T)$ we copy the surrounding rule and replace the head with $\pi_{\texttt{typed}}(e, C)$:

$$\begin{array}{rl}
\pi_{\texttt{typed}}(e_1, C) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(e_1, T_e),\ \ ?\texttt{Fun}(T_e, T_1, T),\ \ \texttt{typed}(e_2, T_1),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(e_2, C) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(e_1, T_e),\ \ ?\texttt{Fun}(T_e, T_1, T),\ \ \texttt{typed}(e_2, T_1),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(b, C') & \texttt{:-}\ \ ?\texttt{lam}(e, x, T_1, b),\ \ !\texttt{bind}(C', C, x, T_1),\ \ \texttt{typed}(b, T_2),\ \ !\texttt{Fun}(T, T_1, T_2),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(e, C) & \texttt{:-}\ \ ?\texttt{main}(p, e),\ \ !\texttt{empty}(C),\ \ \texttt{typed}(e, T).
\end{array}$$

Note how the derived relation finds the co-functional column C of an expression e_1 by querying itself recursively for parent node of e_1. This way, the derived relation retraces the original context propagation. However, this initial version of $\pi_{\texttt{typed}}$ only yields a context for e if e is in `typed`, even though this does not influence which context is returned. To break this dependency, we drop all atoms from $\pi_{\texttt{typed}}$ that do not contribute to determining the co-functional column. We can also simplify `typed`, but we may only remove atoms that are infallible (`!bind`, `!empty`, $\pi_{\texttt{typed}}$) to preserve ill-typed terms. This yields the following minimal rule set with a clear division of labor: $\pi_{\texttt{typed}}$ propagates and extends the context, whereas `typed` does the checking and only mentions the context in the `var` rule.

$$\begin{array}{rl}
\texttt{typed}(e, T) & \texttt{:-}\ \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T). \\
\texttt{typed}(e, T) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(e_1, T_e),\ \ ?\texttt{Fun}(T_e, T_1, T),\ \ \texttt{typed}(e_2, T_1). \\
\texttt{typed}(e, T) & \texttt{:-}\ \ ?\texttt{lam}(e, x, T_1, b),\ \ \texttt{typed}(b, T_2),\ \ !\texttt{Fun}(T, T_1, T_2). \\
\texttt{typed}(e, T) & \texttt{:-}\ \ \pi_{\texttt{typed}}(e, C),\ \ ?\texttt{var}(e, x),\ \ \texttt{lookup}(C, x, T). \\
\texttt{ok}(p) & \texttt{:-}\ \ ?\texttt{main}(p, e),\ \ \texttt{typed}(e, T). \\
\\
\pi_{\texttt{typed}}(e_1, C) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(e_2, C) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(b, C') & \texttt{:-}\ \ ?\texttt{lam}(e, x, T_1, b),\ \ !\texttt{bind}(C', C, x, T_1),\ \ \pi_{\texttt{typed}}(e, C). \\
\pi_{\texttt{typed}}(e, C) & \texttt{:-}\ \ ?\texttt{main}(p, e),\ \ !\texttt{empty}(C).
\end{array}$$

It is easy to show by induction that $\pi_{\texttt{typed}}$ satisfies the functional dependency $e \rightsquigarrow C$. As we explained for Challenge 2 in Section 2, this is sufficient to ensure the finiteness of $\pi_{\texttt{typed}}$. Hence, incremental Datalog solvers can apply their bottom-up evaluation strategy to this Datalog program.

3.3 Formalizing Transformation *CoFunTrans*

We formalize the transformation *CoFunTrans* that we described informally above. The transformation takes a Datalog program as input and rewrites it to utilize co-functional dependencies. The transformation operates in two steps. First, we revise the signatures of existing relations and add the signatures of derived relations π_R. Second, we revise the rules of existing relations and add new rules for derived relations π_R.

CoFunTrans signatures. Let Σ be the set of relational signatures of the input Datalog program. Then the rewritten Datalog program has signatures *CoFunTransSigs*(Σ) defined as follows:

$$\textit{CoFunTrans}\text{-UpdateSig}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \quad \mathit{codepCols} = \{i \mid (I \overset{co}{\rightsquigarrow} i) \in G\} \\ J = \{1, \ldots, n\} \setminus \mathit{codepCols} \quad F' = \mathit{deleteShift}(F, \mathit{codepCols})\end{array}}{(R : T_J \,|\, F', \emptyset) \in \mathit{CoFunTransSigs}(\Sigma)}$$

$$\textit{CoFunTrans}\text{-DeriveSig}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \quad (I \overset{co}{\rightsquigarrow} i) \in G \\ f = \{1, \ldots, |I|\} \rightsquigarrow |I| + 1\end{array}}{(\pi_{R,i} : T_I \times T_i \,|\, \{f\}, \emptyset) \in \mathit{CoFunTransSigs}(\Sigma)}$$

Rule *CoFunTrans*-UpdateSig updates the signatures of existing relations R by dropping all columns $\mathit{codepCols}$ that are co-functionally dependent . The updated signature of R only has columns J of types T_J left. The functional dependencies F are updated accordingly and the co-functional dependencies G are dropped entirely. In particular, function $\mathit{deleteShift}(F, \mathit{codepCols})$ deletes $\mathit{codepCols}$ from the functional dependencies in F and shifts the remaining indices to skip dropped columns. For example, $(\mathtt{typed} : C \times e \times T \,|\, \{\{1, 2\} \rightsquigarrow 3\}, \{\{2\} \overset{co}{\rightsquigarrow} 1\})$ becomes $(\mathtt{typed} : e \times T \,|\, \{\{1\} \rightsquigarrow 2\}, \emptyset)$ after dropping column C.

Rule *CoFunTrans*-DeriveSig generates a separate signature $\pi_{R,i}$ for each co-functional dependency $I \overset{co}{\rightsquigarrow} i$ of a relation R. Relation $\pi_{R,i}$ maps columns I of types T_I to column i of type T_i, as expressed by its functional dependency f.

CoFunTrans rules. Let Σ be the set of relational signatures of the input Datalog program and let P be the set of rules of the input Datalog program. Then the rewritten Datalog program has rules $\mathit{CoFunTransRules}(\Sigma, \mathrm{P})$ defined by the inference rules shown in Figure 2.

$$\textit{CoFunTrans}\text{-DropCodepArgs}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \quad \mathit{codepCols} = \{i \mid (I \overset{co}{\rightsquigarrow} i) \in G\} \\ J = \{1, \ldots, n\} \setminus \mathit{codepCols}\end{array}}{\lfloor \mathtt{R}(\overline{x}) \rfloor = \mathtt{R}(x_J)}$$

$$\textit{CoFunTrans}\text{-UpdateRule}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \quad (\mathtt{R}(\overline{x}) \mathrel{:\!-} a_1, \ldots, a_m.) \in \mathrm{P} \\ \Pi = \{\pi_{\mathtt{R},\mathtt{i}}(x_I, x_i) \mid (I \overset{co}{\rightsquigarrow} i) \in G\}\end{array}}{(\lfloor \mathtt{R}(\overline{x}) \rfloor \mathrel{:\!-} \Pi, \lfloor a_1 \rfloor, \ldots, \lfloor a_m \rfloor.) \in \mathit{CoFunTransRules}(\Sigma, \mathrm{P})}$$

$$\textit{CoFunTrans}\text{-DeriveRule}\ \frac{\begin{array}{cc}(R : T_1 \times \ldots \times T_n \,|\, F_R, G_R) \in \Sigma & (I \overset{co}{\rightsquigarrow} i) \in G_R \\ (\mathtt{Q}(\overline{y}) \mathrel{:\!-} a_1, \ldots, a_k, \mathtt{R}(\overline{x}), a_{k+2}, \ldots, a_l.) \in \mathrm{P} & (Q : U_1 \times \ldots \times U_m \,|\, F_Q, G_Q) \in \Sigma \\ A = \{\lfloor a_1 \rfloor, \ldots, \lfloor a_k \rfloor, \lfloor a_{k+2} \rfloor, \ldots, \lfloor a_l \rfloor\} & \Pi = \{\pi_{\mathtt{Q},\mathtt{j}}(y_J, y_j) \mid (J \overset{co}{\rightsquigarrow} j) \in G_Q\}\end{array}}{(\pi_{\mathtt{R},\mathtt{i}}(x_I, x_i) \mathrel{:\!-} \mathbf{slice}_{x_i}(A \cup \Pi).) \in \mathit{CoFunTransRules}(\Sigma, \mathrm{P})}$$

Fig. 2. Set of inference rules defining $\mathit{CoFunTransRules}(\Sigma, \mathrm{P})$.

Rule *CoFunTrans*-DropCodepArgs defines an auxiliary function $\lfloor a \rfloor$ on atoms that removes co-functionally dependent arguments from calls of R. We use this function in the other two rules.

Rule *CoFunTrans*-UpdateRule updates the Datalog rules of existing relations R by making three changes. First, we remove co-functionally dependent columns from the rule head. Second, we insert queries against the newly derived relations $\pi_{R,i}$ into the body of the rule for each co-functionally dependent column i. Third, we remove co-functionally dependent arguments from calls to other relations in the rule body.

Rule *CoFunTrans*-DeriveRule generates Datalog rules for the new relations $\pi_{R,i}$. Specifically, we generate one Datalog rule for each call of R and each co-functional dependency $I \overset{co}{\rightsquigarrow} i$ of relation R. Suppose the call of R occurs in a rule of Q. We derive the new rule by changing the rule of Q in three ways. First, we exchange the rule head since we are only interested in learning how x_I determines x_i. Second, we adapt the rule body just like *CoFunTrans*-UpdateRule did: remove co-functionally dependent arguments and insert queries $\pi_{Q,j}$. This yields the body. Third, we slice the resulting $(A \cup \Pi)$ to only retain those that contribute to x_i.

The resulting Datalog program witnesses co-functional dependencies through the derived relations $\pi_{R,i}$. Note that the transformation only preserves the semantics of the main relation (e.g. `ok`), but not the semantics of individual Datalog relations. This is intended as we wanted to restrict `typed` to become finite. Importantly, we only remove unnecessary tuples from `typed` such that the main relation is preserved: $p \in$ `ok` if and only if $p \in$ *CoFunTrans*(`ok`).

Correctness. We formulate precisely under which conditions the transformation is correct and preserves the semantics of the transformed type system. Note that we have not worked out a formal proof of correctness, but rather present the key invariants and properties that ensure correctness.

The formalization assumes a denotational semantics $[\![\mathrm{P}]\!]_{Base}$ for Datalog programs P [Alvarez-Picallo et al. 2019]. The denotational semantics takes a set of base facts $Base$ as input (the extensional database in Datalog lingo) and computes the set of facts $Deriv$ derived by P from $Base$ (yielding the intensional database). Both $Base$ and $Deriv$ consist of ground tuples $R(C, \ldots, C)$, where R is a relation name and C are constants. In our work, the base facts describe the user program as we illustrated in Section 2.

Based on the denotational semantics we define the correctness theorem of *CoFunTrans*. Transformation *CoFunTrans* is correct because it leads to less derivable tuples, but all relevant tuples are still derivable. Hence, every program that is typeable with the original type system is typeable with the transformed type system. Without loss of generality, we assume there is a relation *Main* used as the entry point of type checking.

Theorem 1 (Correctness of *CoFunTrans*)**.** Let Σ be the set of relational signatures of the input Datalog program, let P be the set of rules of the input Datalog program, and let *Main* be the entry point of the type system. For all base relations *Base*, $Main(\overline{x}) \in [\![\mathrm{P}]\!]_{Base}$ if and only if $Main(\overline{x}) \in [\![\mathit{CoFunTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.

To prove this theorem, we need to ensure that the derived π relations correctly capture co-functional dependencies $I \overset{co}{\rightsquigarrow} i$: It uniquely maps x_I to x_i. That is, we need to ensure that if the original relation R contains a tuple, the derived π relation contains a tuple representing a projection of the tuple according to the co-functional dependency. In addition, we need to ensure that the derived π relations actually describe functional dependencies to guarantee that Datalog bottom-up evaluation succeeds for the transformed Datalog program.

Lemma 1 (Derived π is correct)**.** Let $(R : T_1 \times \ldots \times T_n | F, G) \in \Sigma$ with $(I \overset{co}{\rightsquigarrow} i) \in G$. For all base relations *Base*, if $R(\overline{x}) \in [\![\mathrm{P}]\!]_{Base}$, then the following propositions hold:

1. $\pi_{R,\mathrm{i}}(x_I, x_i) \in [\![\mathit{CoFunTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.
2. For all $x' \neq x_i$, $\pi_{R,i}(x_I, x') \notin [\![\mathit{CoFunTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.

Lemma 1 is the key property of *CoFunTrans* required to prove correctness.

4 TRANSFORMATION 2: CONTEXT FUSION

Transformation *CoFunTrans* from the previous section makes a Datalog-encoded type system amenable to bottom-up evaluation. It does so by eliminating co-functional dependencies in favor of functional dependencies. For a type system, this means that class tables, typing contexts, and other contextual information is uniquely associated with each expression. While this enabled bottom-up evaluation, it also introduced a new problem: Even a slight change to contextual information will affect all expressions. This is the problem of context propagation we introduced as Challenge 4.

Context propagation is problematic whenever the context consists of compound information (e.g., a typing context). When parts of the context are changed (e.g., the type of some variable), the entire context will regarded as changed. This is because incremental Datalog solvers only trace dependencies between relations and propagate inserted and deleted tuples, but they cannot trace changes to individual components of those tuples. Therefore, when the type of a variable changes, all typing contexts that contain that binding change, and thus all tuples that associate these contexts to expressions need updating.

In this section, we present a Datalog transformation that eliminates intermediate compound data. Specifically, we eliminate context information represented as immutable maps, which is produced by the `!empty` and the `!bind` constructors and consumed with `lookup`. Our rewriting can be regarded as a special case of deforestation [Wadler 1990] for immutable maps but for Datalog programs and with support for recursively defined relations. Note also that immutable maps can encode sets as $\texttt{Map}[A, \text{Unit}]$ and lists as $\texttt{Map}[\texttt{Int}, A]$, such that our rewriting supports many type system specifications. Nonetheless, our primary motivation was the elimination of intermediate typing contexts, which is why we call the transformation "context fusion".

4.1 Context Fusion by Example

Consider again the Datalog rules for the simply typed lambda calculus, as produced by transformation *CoFunTrans* from the previous section.

$$\begin{aligned}
\texttt{typed}(e,T) &\ \texttt{:-}\ \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T).\\
\texttt{typed}(e,T) &\ \texttt{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \texttt{typed}(e_1,T_e),\ \ ?\texttt{Fun}(T_e,T_1,T),\ \ \texttt{typed}(e_2,T_1).\\
\texttt{typed}(e,T) &\ \texttt{:-}\ \ ?\texttt{lam}(e,x,T_1,b),\ \ \texttt{typed}(b,T_2),\ \ !\texttt{Fun}(T,T_1,T_2).\\
\texttt{typed}(e,T) &\ \texttt{:-}\ \ \pi_{\texttt{typed}}(e,C),\ \ ?\texttt{var}(e,x),\ \ \texttt{lookup}(C,x,T).\\
\texttt{ok}(p) &\ \texttt{:-}\ \ ?\texttt{main}(p,e),\ \ \texttt{typed}(e,T).
\end{aligned}$$

$$\begin{aligned}
\pi_{\texttt{typed}}(e_1,C) &\ \texttt{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \pi_{\texttt{typed}}(e,C).\\
\pi_{\texttt{typed}}(e_2,C) &\ \texttt{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \pi_{\texttt{typed}}(e,C).\\
\pi_{\texttt{typed}}(b,C') &\ \texttt{:-}\ \ ?\texttt{lam}(e,x,T_1,b),\ \ !\texttt{bind}(C',C,x,T_1),\ \ \pi_{\texttt{typed}}(e,C).\\
\pi_{\texttt{typed}}(e,C) &\ \texttt{:-}\ \ ?\texttt{main}(p,e),\ \ !\texttt{empty}(C).
\end{aligned}$$

Our goal is to eliminate the typing context produced by $\pi_{\texttt{typed}}$ and consumed by `lookup` in the `var` rule. Though we consider `lookup` to be a built-in operation, it can be defined in Datalog as follows:

$$\begin{aligned}
\texttt{lookup}(m,k,v) &\ \texttt{:-}\ \ ?\texttt{bind}(m,\_,k,v).\\
\texttt{lookup}(m,k,v) &\ \texttt{:-}\ \ ?\texttt{bind}(m,m',k',v'),\ \ k \neq k',\ \ \texttt{lookup}(m',k,v).
\end{aligned}$$

The first rule yields value v if map m starts with a binding for key k. The second rule continues lookup in the rest of the map m' if k differs from k'.

To eliminate the context, we want to fuse $\pi_{\texttt{typed}}$ and `lookup`. Specifically, since relation $\pi_{\texttt{typed}}$ has a functional dependency $e \rightsquigarrow C$, it uniquely associates a context to an expression. Thus, instead of performing `lookup` on the context, can't we derive a specialized `lookup` relation that operates on the expression directly? Indeed, this is what our second transformation does.

We derive a specialized `lookup` relation $\varphi_{\texttt{typed}} : e \times v \times T$ that *finds* the binding of a variable v given an expression e. We find bindings by mimicking the rules of $\pi_{\texttt{typed}}$, When a `!bind` occurs in $\pi_{\texttt{typed}}$, we inline the definition of `lookup` to check if we have found the desired entry. For the simply typed lambda calculus we obtain the following rules:

$$
\begin{array}{rcl}
\dots & & \dots \\
\texttt{typed}(e, T) & \texttt{:-} & ?\texttt{var}(e, x),\ \ \varphi_{\texttt{typed}}(e, x, T). \\
\\
\varphi_{\texttt{typed}}(e_1, k, v) & \texttt{:-} & ?\texttt{app}(e, e_1, e_2),\ \ \varphi_{\texttt{typed}}(e, k, v). \\
\varphi_{\texttt{typed}}(e_2, k, v) & \texttt{:-} & ?\texttt{app}(e, e_1, e_2),\ \ \varphi_{\texttt{typed}}(e, k, v). \\
\varphi_{\texttt{typed}}(b, k, v) & \texttt{:-} & ?\texttt{lam}(e, x, T_1, b),\ \ k = x,\ \ v = T_1. \\
\varphi_{\texttt{typed}}(b, k, v) & \texttt{:-} & ?\texttt{lam}(e, x, T_1, b),\ \ k \neq x,\ \ \varphi_{\texttt{typed}}(e, k, v).
\end{array}
$$

For applications, $\pi_{\texttt{typed}}$ propagated the context of the parent term e. Hence, $\varphi_{\texttt{typed}}$ continues its search for k in the parent term. For lambdas, $\pi_{\texttt{typed}}$ yielded an extended context $!\texttt{bind}(C', C, x, T_1)$. We inline the definition of `lookup` and hence obtain two $\varphi_{\texttt{typed}}$ rules. First, we yield T_1 if the bound variable x is the entry k we are looking for. Second, we continue searching in the parent term if x and k differ. For the main program, $\pi_{\texttt{typed}}$ yields the empty context $!\texttt{empty}(C)$. Since lookup fails on the empty context, we do not add a rule to $\varphi_{\texttt{typed}}$. Consequently, $\varphi_{\texttt{typed}}$ will fail (as it should) when we reached the root node and have not found a binding.

4.2 Formalizing Transformation *CtxFusionTrans*

We formalize the transformation *CtxFusionTrans* that we exemplified above. The transformation takes a Datalog program as input and rewrites it to derive and apply find relations φ_R. We first derive the new signatures and then update and add rules to the Datalog program.

CtxFusionTrans signatures. Let Σ be the set of relational signatures of the input Datalog program. The rewritten Datalog program has signatures *CtxFusionTransSigs*(Σ) defined as follows:

$$
\mathit{CtxFusionTrans}\text{-DeriveSig}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \\ (I \rightsquigarrow i) \in F \quad T_i = \texttt{Map}[K, V] \\ f = \{1, \ldots, |I| + 1\} \rightsquigarrow |I| + 2\end{array}}{(\varphi_{R,i} : T_I \times K \times V \,|\, \{f\}, \emptyset) \in \mathit{CtxFusionTransSigs}(\Sigma)}
$$

$$
\mathit{CtxFusionTrans}\text{-RetainSig}\ \frac{(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma}{(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \mathit{CtxFusionTransSigs}(\Sigma)}
$$

Rule *CtxFusionTrans*-DeriveSig generates a signature for the find relations φ_R. We generate a separate find relation for each functional dependency $I \rightsquigarrow i$ where column i has a `Map` type. That is, whenever it is possible to uniquely determine a map from other columns I, we want to find bindings based on I. The find relation uniquely maps values of types T_I together with a key of type K to a value of type V, as expressed by the functional dependency f.

Rule *CtxFusionTrans*-RetainSig merely retains all existing signatures. Usually, it is possible to drop relations that only produce a map since we won't need them after the transformation. For example, we dropped relation $\pi_{\texttt{typed}}$ in our example from above, using $\varphi_{\texttt{typed}}$ instead. However, our transformation does not account for this simple post-processing.

CtxFusionTrans rules. Let Σ be the set of relational signatures of the input Datalog program and let P be the set of rules of the input Datalog program. Then the rewritten Datalog program has rules *CtxFusionTransRules*(Σ, P) defined by the inference rules shown in Figure 3. In computing *CtxFusionTransRules*(Σ, P), we construct intermediate sets $\mathit{Step}_z(\Sigma, \mathrm{P})$ that contain rules after a z-fold unfolding of the $!\texttt{bind}$ constructor. Note that the unfolding is bounded by the number of syntactic occurrences of $!\texttt{bind}$ in the original rules.

$$\textit{CtxFusionTrans}\text{-Init}\;\frac{\begin{array}{cc}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma & (\mathtt{R}(\overline{x}) \,\text{:-}\, a_1, \ldots, a_m.) \in \mathrm{P} \\ (I \rightsquigarrow i) \in F & T_i = \mathtt{Map}[K, V]\end{array}}{(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, \mathtt{lookup}(x_i, k, v).) \in \mathit{Step}_0(\Sigma, \mathrm{P})}$$

$$\textit{CtxFusionTrans}\text{-Unfold}\;\frac{\begin{array}{c}(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, \mathtt{lookup}(x_i, k, v).) \in \mathit{Step}_z(\Sigma, \mathrm{P}) \\ !\mathtt{bind}(M, M', k', v') \in \{a_1, \ldots, a_m\} \quad a_1, \ldots, a_m \vdash x_i = M\end{array}}{(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, k \neq k', \mathtt{lookup}(M', k, v).) \in \mathit{Step}_{z+1}(\Sigma, \mathrm{P})}$$

$$\textit{CtxFusionTrans}\text{-Bound}\;\frac{\begin{array}{c}(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, \mathtt{lookup}(x_i, k, v).) \in \mathit{Step}_z(\Sigma, \mathrm{P}) \\ !\mathtt{bind}(M, M', k', v') \in \{a_1, \ldots, a_m\} \quad a_1, \ldots, a_m \vdash x_i = M\end{array}}{(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, k = k', v = v'.) \in \mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})}$$

$$\textit{CtxFusionTrans}\text{-Delegate}\;\frac{\begin{array}{cc}\multicolumn{2}{c}{(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, \mathtt{lookup}(x_i, k, v).) \in \mathit{Step}_z(\Sigma, \mathrm{P})} \\ \mathtt{Q}(\overline{y}) \in \{a_1, \ldots, a_m\} & a_1, \ldots, a_m \vdash x_i = y_j \\ (Q : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma & (J \rightsquigarrow j) \in F\end{array}}{(\varphi_{\mathtt{R},\mathtt{i}}(x_I, k, v) \,\text{:-}\, a_1, \ldots, a_m, \varphi_{\mathtt{Q},\mathtt{j}}(y_J, k, v).) \in \mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})}$$

$$\textit{CtxFusionTrans}\text{-Replace}\;\frac{\begin{array}{cc}\multicolumn{2}{c}{(\mathtt{R}(\overline{x}) \,\text{:-}\, a_1, \ldots, a_i, \mathtt{lookup}(M, k, v), a_{i+2}, \ldots, a_m.) \in \mathrm{P}} \\ \mathtt{Q}(\overline{y}) \in \{a_1, \ldots, a_m\} & a_1, \ldots, a_m \vdash M = y_j \\ (Q : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma & (J \rightsquigarrow j) \in F\end{array}}{(\mathtt{R}(\overline{x}) \,\text{:-}\, a_1, \ldots, a_i, \varphi_{\mathtt{Q},\mathtt{j}}(y_J, k, v), a_{i+2}, \ldots, a_m.) \in \mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})}$$

$$\textit{CtxFusionTrans}\text{-Retain}\;\frac{\begin{array}{c}(\mathtt{R}(\overline{x}) \,\text{:-}\, a_1, \ldots, a_m.) \in \mathrm{P} \\ \textit{CtxFusionTrans}\text{-Replace not applicable}\end{array}}{(\mathtt{R}(\overline{x}) \,\text{:-}\, a_1, \ldots, a_m.) \in \mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})}$$

Fig. 3. Set of inference rules defining *CtxFusionTransRules*(Σ, P).

Rule *CtxFusionTrans*-Init derives the initial $\varphi_{R,i}$ rule for any R that has a functional dependency $I \rightsquigarrow i$ with column i being a `Map`. Given x_I and k, the initial rule uses $a_1, \ldots, a_m$ to uniquely obtain x_i and then perform a `lookup` on that. In the subsequent rules, we try to eliminate the invocation of `lookup` and with it the need for obtaining the map x_i explicitly.

Rules *CtxFusionTrans*-Unfold and *CtxFusionTrans*-Bound have the same premises. They check if `lookup` is invoked on an explicitly constructed map. To this end, we check if any of the atoms $a_1, \ldots, a_m$ is an invocation of !`bind` and if the !`bind`-constructed map M is used in `lookup`. We write $a_1, \ldots, a_m \vdash x_i = M$ to mean that x_i and M unify to the same logic variable under $a_1, \ldots, a_m$, which is decidable in Datalog. If so, we know that the `lookup` occurs on top of M. We can thus inline `lookup`. Rule *CtxFusionTrans*-Unfold captures the case where $k \neq k'$ and `lookup` thus must continue on the rest of the map M'. Since the resulting rule still contains `lookup`, we add the rule to $\mathit{Step}_{z+1}(\Sigma, \mathrm{P})$ to allow further transformation. Rule *CtxFusionTrans*-Bound captures the case where $k = k'$, so that we can yield $v = v'$. Since *CtxFusionTrans*-Bound fully eliminated the `lookup` call, we add the resulting rule to the output of the transformation.

Rule *CtxFusionTrans*-Delegate checks if `lookup` is invoked on a context obtained from another relation. That is the case if any of the atoms $a_1, \ldots, a_m$ is a query $\mathtt{Q}(\overline{y})$ and the map x_i corresponds to y_j for some j. Now, if Q has a functional dependency $J \rightsquigarrow j$ and uniquely determines the map y_j, then we can use the $\varphi_{Q,j}$ relation we created in *CtxFusionTrans*-DeriveSig for Q. That is, we delegate the search in R and continue searching in Q, which may lead to a (mutually) recursively defined search relation.

Rules *CtxFusionTrans*-Replace and *CtxFusionTrans*-Retain propagate the original rules from P. Rule *CtxFusionTrans*-Replace applies to rules that contain a `lookup` on a map that is uniquely obtained from relation Q. We replace these lookups by the corresponding search $\varphi_{Q,j}$. Rule *CtxFusionTrans*-Retain copies over all other rules unchanged.

Note that rules can starve within Step_z and never make it to *CtxFusionTransRules*. This is intended and accounts for the cases where the original `lookup` would have failed as well. In particular, a lookup on an empty map will not result in a *CtxFusionTransRules* rule.

Correctness. We formulate a correctness theorem for *CtxFusionTrans*. This transformation is an optimization and does not affect the derivable tuples in any way. Hence, every tuple that is derivable by the original type system is derivable by the transformed type system.

Theorem 2 (Correctness of *CtxFusionTrans*)**.** Let Σ be the relational signatures of the input Datalog program, let P be the rules of the input Datalog program, and let $(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma$. For all base relations *Base*, $R(\overline{x}) \in [\![\mathrm{P}]\!]_{Base}$ if and only if $R(\overline{x}) \in [\![\mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.

The only rewriting that changes existing relations is *CtxFusionTrans*-Replace, which replaces invocations of `lookup` with invocations of φ. To justify this rewriting, we need to ensure the derived φ relations are correct, that is, they resolve key k to the same value v as `lookup`. We formulate the following lemma which ensures we can actually replace `lookup` with the derived φ relations. It states that `lookup` will resolve key k to the same value v within context x_j as the newly derived relation $\varphi_{Q,j}$. Note that Q functionally determines the context x_j based on x_J which in turn will be source arguments of $\varphi_{Q,j}$.

Lemma 2 (Derived φ is correct)**.** Let $(Q : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma$ with $(J \rightsquigarrow j) \in F$. For all base relations *Base*, if $Q(x_J, x_j) \in [\![\mathrm{P}]\!]_{Base}$ and $\texttt{lookup}(x_j, k, v) \in [\![\mathrm{P}]\!]_{Base}$, then $\varphi_{Q,\mathtt{j}}(x_J, k, v) \in [\![\mathit{CtxFusionTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.

This characterizes under which conditions the correctness of *CtxFusionTrans* is ensured.

4.3 Example Revisited

We illustrate the step-wise application of *CtxFusionTransRules* to the relevant rules of the simply typed lambda calculus.

Input rules:

$\pi_{\texttt{typed}}(e_1, C)$:- $?\texttt{app}(e, e_1, e_2),\ \pi_{\texttt{typed}}(e, C).$
$\pi_{\texttt{typed}}(e_2, C)$:- $?\texttt{app}(e, e_1, e_2),\ \pi_{\texttt{typed}}(e, C).$
$\pi_{\texttt{typed}}(b, C')$:- $?\texttt{lam}(e, x, T_1, b),\ !\texttt{bind}(C', C, x, T_1),\ \pi_{\texttt{typed}}(e, C).$
$\pi_{\texttt{typed}}(e, C)$:- $?\texttt{main}(p, e),\ !\texttt{empty}(C).$
$\texttt{typed}(e, T)$:- $\pi_{\texttt{typed}}(e, C),\ ?\texttt{var}(e, x),\ \texttt{lookup}(C, x, T).$

$\mathit{Step}_0(\Sigma, \mathrm{P})$:

$\varphi_{\texttt{typed}}(e_1, k, v)$:- $?\texttt{app}(e, e_1, e_2),\ \pi_{\texttt{typed}}(e, C),\ \texttt{lookup}(C, k, v).$
$\varphi_{\texttt{typed}}(e_2, k, v)$:- $?\texttt{app}(e, e_1, e_2),\ \pi_{\texttt{typed}}(e, C),\ \texttt{lookup}(C, k, v).$
$\varphi_{\texttt{typed}}(b, k, v)$:- $?\texttt{lam}(e, x, T_1, b),\ !\texttt{bind}(C', C, x, T_1),\ \pi_{\texttt{typed}}(e, C),\ \texttt{lookup}(C', k, v).$
$\varphi_{\texttt{typed}}(e, k, v)$:- $?\texttt{main}(p, e),\ !\texttt{empty}(C),\ \texttt{lookup}(C, k, v).$

$\mathit{Step}_1(\Sigma, \mathrm{P})$:

$\varphi_{\texttt{typed}}(b, k, v)$:- $?\texttt{lam}(e, x, T_1, b),\ !\texttt{bind}(C', C, x, T_1),\ \pi_{\texttt{typed}}(e, C),\ k \neq x,\ \texttt{lookup}(C, k, v).$

CtxFusionTransRules(Σ, P):

$$\begin{aligned}
\varphi_{\texttt{typed}}(e_1, k, v) &\ \text{:-}\ \texttt{?app}(e, e_1, e_2),\ \ \pi_{\texttt{typed}}(e, C),\ \ \varphi_{\texttt{typed}}(e, k, v).\\
\varphi_{\texttt{typed}}(e_2, k, v) &\ \text{:-}\ \texttt{?app}(e, e_1, e_2),\ \ \pi_{\texttt{typed}}(e, C),\ \ \varphi_{\texttt{typed}}(e, k, v).\\
\varphi_{\texttt{typed}}(b, k, v) &\ \text{:-}\ \texttt{?lam}(e, x, T_1, b),\ \ \texttt{!bind}(C', C, x, T_1),\ \ \pi_{\texttt{typed}}(e, C),\ \ k = x,\ \ v = T_1.\\
\varphi_{\texttt{typed}}(b, k, v) &\ \text{:-}\ \texttt{?lam}(e, x, T_1, b),\ \ \texttt{!bind}(C', C, x, T_1),\ \ \pi_{\texttt{typed}}(e, C),\ \ k \neq x,\ \ \varphi_{\texttt{typed}}(e, k, v).\\
\texttt{typed}(e, T) &\ \text{:-}\ \pi_{\texttt{typed}}(e, C),\ \ \texttt{?var}(e, x),\ \ \varphi_{\texttt{typed}}(e, x, T).
\end{aligned}$$

A subsequent optimization of the derived rules will remove all invocations of $\pi_{\texttt{typed}}$ and `!bind`. This is supported by our implementation and will yield exactly those rules shown in Section 4.1.

4.4 Optimizing Search Relations φ_R

Our transformation *CtxFusionTrans* successfully eliminated all intermediate contexts and introduced a bottom-up find function instead. As we will show in our empirical evaluation, the resulting Datalog code yields far superior incremental performance. However, there is one issue we need to take care of first: The derived find relations φ_R enumerate all referable bindings, not just those required by actual references.

Consider the example term $\lambda x{:}\mathrm{Unit}.\,(1+2)+(3+4)$, where we used additions and numeric literals for convenience. Although this program contains no variable references, $\varphi_{\texttt{typed}}$ contains all of the entries shown in the table on the right. That is, $\varphi_{\texttt{typed}}$ contains one entry for each variable and each expression where that variable is in scope. This does not scale very well and it is unneeded.

$\varphi_{\texttt{typed}}$	e	x	T
	1	x	Unit
	2	x	Unit
	$1+2$	x	Unit
	3	x	Unit
	4	x	Unit
	$3+4$	x	Unit
	$(1+2)+(3+4)$	x	Unit

Indeed it is sufficient to consider variables that are being referenced in an expression. Fortunately, we can derive an optimized version of $\varphi_{\texttt{typed}}$ by restricting its entries. Specifically, we implemented a simple magic set transformation [Beeri and Ramakrishnan 1991] to derive a helper relation ρ_R that restricts φ_R to those tuples for which a lookup is needed. In particular, when φ_R corresponds to variable lookup, ρ_R corresponds to the free variables of an expression. We restrict φ_R to those tuples that ρ_R considers relevant:

$$\begin{aligned}
\varphi_{\texttt{typed}}(e_1, k, v) &\ \text{:-}\ \rho_{\texttt{typed}}(e_1, k),\ \ \texttt{?app}(e, e_1, e_2),\ \ \varphi_{\texttt{typed}}(e, k, v).\\
\varphi_{\texttt{typed}}(e_2, k, v) &\ \text{:-}\ \rho_{\texttt{typed}}(e_2, k),\ \ \texttt{?app}(e, e_1, e_2),\ \ \varphi_{\texttt{typed}}(e, k, v).\\
\varphi_{\texttt{typed}}(b, k, v) &\ \text{:-}\ \rho_{\texttt{typed}}(b, k),\ \ \texttt{?lam}(e, x, T_1, b),\ \ k = x,\ \ v = T_1.\\
\varphi_{\texttt{typed}}(b, k, v) &\ \text{:-}\ \rho_{\texttt{typed}}(b, k),\ \ \texttt{?lam}(e, x, T_1, b),\ \ k \neq x,\ \ \varphi_{\texttt{typed}}(e, k, v).
\end{aligned}$$

$$\begin{aligned}
\rho_{\texttt{typed}}(e, x) &\ \text{:-}\ \texttt{?var}(e, x).\\
\rho_{\texttt{typed}}(e, k) &\ \text{:-}\ \texttt{?app}(e, e_1, e_2),\ \ \rho_{\texttt{typed}}(e_1, k).\\
\rho_{\texttt{typed}}(e, k) &\ \text{:-}\ \texttt{?app}(e, e_1, e_2),\ \ \rho_{\texttt{typed}}(e_2, k).\\
\rho_{\texttt{typed}}(e, k) &\ \text{:-}\ \texttt{?lam}(e, x, T_1, b),\ \ k \neq x,\ \ \rho_{\texttt{typed}}(b, k).
\end{aligned}$$

For $\lambda x{:}\mathrm{Unit}.\,(1+2)+(3+4)$, relation $\rho_{\texttt{typed}}$ remains empty since no free variables occur. Consequently, $\varphi_{\texttt{typed}}$ is empty as well. Our implementation supports this optimization.

5 TRANSFORMATION 3: COLLECTING ERRORS

The traditional formulation of type systems is focused on deciding if a term is well-typed or ill-typed: There either exists a typing derivation or not. However, applications of type systems need more detailed information, namely the reason(s) a typing derivation could not be constructed. In this section, we propose an alternative formulation of type systems that separates finding a term's type from reporting type errors. This allows us (i) to sometimes find a term's type even though there are type errors and (ii) to report multiple type errors for the same term. We present a Datalog transformation that automatically transforms a traditional type system into one with separate error collection. Our transformation is compatible with the previous two transformations from Sections 3 and 4, but it does not require them and can be used independently.

5.1 Collecting Errors by Example

We illustrate how our transformation works by considering the simply typed lambda calculus again. However, to showcase that our transformation can be used independently from the other two transformations, we start with the original type rules from Section 2:

$$\begin{array}{rl}
\texttt{typed}(C,e,T) & \text{:-}\ \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T).\\
\texttt{typed}(C,e,T) & \text{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \texttt{typed}(C,e_1,T_e),\ \ ?\texttt{Fun}(T_e,T_1,T),\ \ \texttt{typed}(C,e_2,T_1).\\
\texttt{typed}(C,e,T) & \text{:-}\ \ ?\texttt{lam}(e,x,T_1,b),\ \ !\texttt{bind}(C',C,x,T_1),\ \ \texttt{typed}(C',b,T_2),\ \ !\texttt{Fun}(T,T_1,T_2).\\
\texttt{typed}(C,e,T) & \text{:-}\ \ ?\texttt{var}(e,x),\ \ \texttt{lookup}(C,x,T).\\
\texttt{ok}(p) & \text{:-}\ \ ?\texttt{main}(p,e),\ \ !\texttt{empty}(C),\ \ \texttt{typed}(C,e,T).
\end{array}$$

The construction of a typing derivation fails when premises are unsatisfiable. However, different premises have different purposes and require different error handling. Therefore, we categorize premises as follows:

- **ReportStuck** is the set of relations whose stuckness should result in a type error. We only track the premises occurring in rules of **ReportStuck** relations. For our example, **ReportStuck** = $\{\texttt{typed}, \texttt{ok}\}$.
- For every $R \in$ **ReportStuck**, **IgnoreStuck**$_R$ is the set of relations that should be ignored when they occur as premises. We use **IgnoreStuck** for those constraints that merely help select the right type rule. For our example, **IgnoreStuck**$_{\texttt{typed}}$ = $\{?\texttt{unit}, ?\texttt{app}, ?\texttt{lam}, ?\texttt{var}\}$ and **IgnoreStuck**$_{\texttt{ok}}$ = $\{?\texttt{main}\}$.
- Some premises $\texttt{R}(\overline{x})$ are known to be infallible and can be ignored during error handling. In our example, $!\texttt{Unity}(T)$ amongst others will never fail and thus cannot produce a type error.

Based on this categorization, we can systematically derive relations $\varepsilon_{\texttt{typed}} : C \times e \times$ **Error** and $\varepsilon_{\texttt{ok}} : p \times$ **Error** that collect the errors that can occur during type checking:

$$\begin{array}{rl}
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \varepsilon_{\texttt{typed}}(C,e_1,\mathit{err}).\\
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \texttt{typed}(C,e_1,T_e),\ \ \neg ?\texttt{Fun}(T_e,T_1,T),\ \ \mathit{err} = \text{“\textit{expected Fun type}”}\\
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \varepsilon_{\texttt{typed}}(C,e_2,\mathit{err}).\\
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{app}(e,e_1,e_2),\ \ \texttt{typed}(C,e_1,T_e),\ \ ?\texttt{Fun}(T_e,T_1,T),\ \ \texttt{typed}(C,e_2,T_2),\\
 & \phantom{\text{:-}}\ \ T_1 \neq T_2,\ \ \mathit{err} = \text{“\textit{type mismatch}”}.\\
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{lam}(e,x,T_1,b),\ \ !\texttt{bind}(C',C,x,T_1),\ \ \varepsilon_{\texttt{typed}}(C',b,\mathit{err}).\\
\varepsilon_{\texttt{typed}}(C,e,\mathit{err}) & \text{:-}\ \ ?\texttt{var}(e,x),\ \ \neg\texttt{lookup}(C,x,T),\ \ \mathit{err} = \text{“\textit{lookup failed}”}.\\
\varepsilon_{\texttt{ok}}(p,\mathit{err}) & \text{:-}\ \ ?\texttt{main}(p,e),\ \ !\texttt{empty}(C),\ \ \varepsilon_{\texttt{typed}}(C,e,\mathit{err}).
\end{array}$$

There is no rule for `unit` because its first premise is in **IgnoreStuck**$_{\texttt{typed}}$ and its second premise is infallible. For `app` we obtain three rules. First, if there are type errors in e_1, we propagate those. We stripped most other premises because they are irrelevant for the recursive call $\varepsilon_{\texttt{typed}}(C, e_1, \mathit{err})$. Second, if the type T_e of e_1 is not a function type (note the negation ¬ in front of ?`Fun`), we generate a new error. Third, we propagate the type errors of e_2. A `lam` cannot introduce a new error and only propagate type errors from the lambda's body. For `var` we obtain a new type error when the `lookup` fails (again note the negation ¬).

Note that the collected type errors are not unique; an expression can have multiple errors. For example, both subterms of an `app` expression can propagate type errors. The derived ε_R relations collect *all* type errors that occur in the program, which was one of our declared goals.

Our other goal was to find a type despite type errors when possible. To this end, we refine what it means for a term to be well-typed in our encoding: A term is well-typed if we can find its type and there is no type error for it. That is, rather than only requiring $(C, e, T) \in \texttt{typed}$, we additionally require $(C, e, err) \notin \varepsilon_{\texttt{typed}}$ for any err. This allows us to retain tuples in `typed` even when an expression contains type errors. Our transformation exploits this to relax the rules of `typed`: Premises that merely perform a check are discarded. For example, our transformation removes the check on an `app`'s argument e_2 and on the body of `main`. The other rules are unaffected:

$$
\begin{array}{rl}
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{unit}(e),\ \ !\texttt{Unit}(T). \\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{app}(e, e_1, e_2),\ \ \texttt{typed}(C, e_1, T_e),\ \ ?\texttt{Fun}(T_e, T_1, T). \\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{lam}(e, x, T_1, b),\ \ !\texttt{bind}(C', C, x, T_1),\ \ \texttt{typed}(C', b, T_2),\ \ !\texttt{Fun}(T, T_1, T_2). \\
\texttt{typed}(C, e, T) & \texttt{:-}\ \ ?\texttt{var}(e, x),\ \ \texttt{lookup}(C, x, T). \\
\texttt{ok}(p) & \texttt{:-}\ \ ?\texttt{main}(p, e).
\end{array}
$$

5.2 Formalizing Transformation *CollectErrorsTrans*

We formalize the transformation *CollectErrorsTrans* that we exemplified above. The transformation takes a Datalog program as input and rewrites it to generate relations ε_R and to relax existing relations. The transformation is parametric in the sets **ReportStuck** and $\textbf{IgnoreStuck}_R$ as described above. We first derive the new signatures and then update and add rules to the Datalog program.

CollectErrorsTrans signatures. Let Σ be the set of relational signatures of the input Datalog program. The rewritten Datalog program has signatures *CollectErrorsTrans*(Σ) defined as follows:

$$
\textit{CollectErrorsTrans}\text{-DeriveSig}\ \frac{\begin{array}{c}(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma \quad R \in \textbf{ReportStuck} \\ J = \{1, \ldots, n\} \setminus \{i \mid (I \rightsquigarrow i) \in F\}\end{array}}{(\varepsilon_R : T_J \times \textbf{Error} \,|\, \emptyset, \emptyset) \in \textit{CollectErrorsTransSigs}(\Sigma)}
$$

$$
\textit{CollectErrorsTrans}\text{-RetainSig}\ \frac{(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \Sigma}{(R : T_1 \times \ldots \times T_n \,|\, F, G) \in \textit{CollectErrorsTransSigs}(\Sigma)}
$$

The first transformation rule adds new signatures ε_R for those relations R that are in **ReportStuck**. The new relation has all columns of R except for those that are functionally dependent. For an algorithmic type system this means that the error relation does not track the computed type. In addition to the columns of R, the error relation ε_R has a new column of type **Error**. A tuple $(t_1, \ldots, t_n, err) \in \varepsilon_R$ means that R is stuck for $(t_1, \ldots, t_n)$. The second transformation rule retains the signatures of all existing relations.

CollectErrorsTrans rules. Let Σ be the set of relational signatures of the input Datalog program and let P be the set of rules of the input Datalog program. Then the rewritten Datalog program has rules *CollectErrorsTransRules*(Σ, P) defined by the inference rules seen in Figure 3.

Rule *CollectErrorsTrans*-Propagate generates a Datalog rule that propagates errors from subderivations upwards. Given the rule of a relation $R \in$ **ReportStuck**, if R calls another relation $Q \in$ **ReportStuck**, then we want to forward the errors of Q. Thus, we generate a rule for ε_R that forwards error err obtained from ε_Q. Since we dropped functionally dependent columns from error relations, we select the appropriate variables x_J and y_K to call ε_R and ε_Q respectively. Finally, we copy a slice of the other atoms A to the resulting rule, namely those that contribute to the call of

$$\textit{CollectErrorsTrans}\text{-Propagate}\ \dfrac{\begin{array}{cc} (\mathtt{R}(\overline{x}) \coloneq a_1, \ldots, a_k, \mathtt{Q}(\overline{y}), a_{k+2}, \ldots, a_m.) \in \mathrm{P} & A = \{a_1, \ldots, a_k, a_{k+2}, \ldots, a_m\} \\ R \in \textbf{ReportStuck} & Q \in \textbf{ReportStuck} \\ (R : T_1 \times \ldots \times T_n \,|\, F_R, G_R) \in \Sigma & J = \{1, \ldots, n\} \setminus \{i \mid (I \rightsquigarrow i) \in F_R\} \\ (Q : T_1 \times \ldots \times T_l \,|\, F_Q, G_Q) \in \Sigma & K = \{1, \ldots, l\} \setminus \{i \mid (I \rightsquigarrow i) \in F_Q\} \end{array}}{(\varepsilon_{\mathtt{R}}(x_J, \mathit{err}) \coloneq \textbf{slice}_{\,y_K}(A), \varepsilon_{\mathtt{Q}}(y_K, \mathit{err}).) \in \textit{CollectErrorsTrans}(\Sigma, \mathrm{P})}$$

$$\textit{CollectErrorsTrans}\text{-NewError}\ \dfrac{\begin{array}{cc} (\mathtt{R}(\overline{x}) \coloneq a_1, \ldots, a_k, \mathtt{Q}(\overline{y}), a_{k+2}, \ldots, a_m.) \in \mathrm{P} & A = \{a_1, \ldots, a_k, a_{k+2}, \ldots, a_m\} \\ R \in \textbf{ReportStuck} & Q \notin \textbf{ReportStuck} \\ Q \notin \textbf{IgnoreStuck}_R & \mathtt{Q}(\overline{y}) \text{ is fallible} \\ (R : T_1 \times \ldots \times T_n \,|\, F_R, G_R) \in \Sigma & J = \{1, \ldots, n\} \setminus \{i \mid (I \rightsquigarrow i) \in F_R\} \\ \multicolumn{2}{c}{\mathit{err} = \text{new error describing the reason } \mathtt{Q}(\overline{y}) \text{ got stuck}} \end{array}}{(\varepsilon_{\mathtt{R}}(x_J, \mathit{err}) \coloneq \textbf{slice}_{\,\overline{y}}(A), \neg\mathtt{Q}(\overline{y}).) \in \textit{CollectErrorsTrans}(\Sigma, \mathrm{P})}$$

$$\textit{CollectErrorsTrans}\text{-RetainSliced}\ \dfrac{\begin{array}{cc} (\mathtt{R}(\overline{x}) \coloneq a_1, \ldots, a_m.) \in \mathrm{P} & R \in \textbf{ReportStuck} \\ \multicolumn{2}{c}{A = \{Q(\overline{y}) \ \mid\ Q(\overline{y}) \in \{a_1, \ldots, a_m\}, Q \in \textbf{IgnoreStuck}_R\}} \end{array}}{(\mathtt{R}(\overline{x}) \coloneq A, \textbf{slice}_{\,\overline{x}}(\{a_1, \ldots, a_m\} \setminus A).) \in \textit{CollectErrorsTrans}(\Sigma, \mathrm{P})}$$

$$\textit{CollectErrorsTrans}\text{-RetainNormal}\ \dfrac{(\mathtt{R}(\overline{x}) \coloneq a_1, \ldots, a_m.) \in \mathrm{P} \quad R \notin \textbf{ReportStuck}}{(\mathtt{R}(\overline{x}) \coloneq a_1, \ldots, a_m.) \in \textit{CollectErrorsTrans}(\Sigma, \mathrm{P})}$$

Fig. 3. Set of inference rules defining $\mathit{CollectErrorsTransRules}(\Sigma, \mathrm{P})$.

ε_Q. This slicing is important for correctness. For example, consider a type rule for binary addition $e_1 + e_2$. Without slicing, we would get the following error rules amongst others:

$$\begin{array}{l} \varepsilon_{\mathtt{typed}}(C, e, \mathit{err}) \coloneq \mathtt{?add}(e, e_1, e_2),\ \ \varepsilon_{\mathtt{typed}}(C, e_1, \mathit{err}),\ \ \mathtt{?Nat}(T_1),\ \ \mathtt{typed}(C, e_2, T_2),\ \ \mathtt{?Nat}(T_2). \\ \varepsilon_{\mathtt{typed}}(C, e, \mathit{err}) \coloneq \mathtt{?add}(e, e_1, e_2),\ \ \mathtt{typed}(C, e_1, T_1),\ \ \mathtt{?Nat}(T_1),\ \ \varepsilon_{\mathtt{typed}}(C, e_2, \mathit{err}),\ \ \mathtt{?Nat}(T_2). \end{array}$$

These rules work fine if one of the operands is ill-typed. But if both operands are ill-typed at the same time, neither rule can fire because of the remaining `typed` constraint on the other operand. Slicing eliminates this problem by discarding those premises that do not help to discover the propagated error.

The second transformation rule *CollectErrorsTrans*-NewError generates error rules for the origin of a stuck premise. If a relation $R \in \textbf{ReportStuck}$ calls another relation $Q \notin \textbf{ReportStuck}$ that is fallible and should not be ignored, then we derive a corresponding error rule. The derived error rule yields a new error description err if $\neg\mathtt{Q}(\overline{y})$, that is, the premise on Q fails. Like in the previous transformation rule, we use slicing to ensure the error rule can fire.

Transformation rule *CollectErrorsTrans*-RetainSliced carries out the relaxation of the original rules for $R \in \textbf{ReportStuck}$. Once again we use slicing, this time to drop premises a_i that do not contribute to discovering the derivable tuples of R . However, the premises A that were ignored by the error rules may never be relaxed. Transformation rule *CollectErrorsTrans*-RetainNormal retains all other Datalog rules unchanged.

Correctness. We formulate a correctness theorem for transformation *CollectErrorsTrans*. This transformation relaxes existing relations and derives error relations to collect stuck derivations instead. That is, all previously derivable tuples are still derivable. In addition, if the original Datalog program will derive a tuple for a relation R marked to report type errors, then there will be no tuple in the corresponding error relation ε_R of the transformed Datalog program.

Theorem 3 (Correctness of *CollectErrorsTrans*)**.** Let Σ be the set of relational signatures of the input Datalog program and P be the set of rules of the input Datalog program. Let $(R : T_1 \times \ldots \times T_n | \, F, G) \in \Sigma$ and $J = \{1, \ldots, n\} \setminus \{i \mid (I \rightsquigarrow i) \in F\}$. For all base relations *Base*, $R(\overline{x}) \in [\![\mathrm{P}]\!]_{Base}$ if and only if the two following properties hold:

(1) $R(\overline{x}) \in [\![\mathit{CollectErrorsTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.
(2) $R \in$ **ReportStuck** implies for all err, $\varepsilon_R(x_J, err) \notin [\![\mathit{CollectErrorsTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$.

To prove this theorem, we need to ensure the error relation ε_R only contains errors for tuples that were previously underivable.

Lemma 3 (Derived ε is correct)**.** Let $(R : T_1 \times \ldots \times T_n | \, F, G) \in \Sigma$ and $J = \{1, \ldots, n\} \setminus \{i | (I \rightsquigarrow i) \in F\}$. For all base relations *Base*, if $\varepsilon_R(\overline{x}, err) \in [\![\mathit{CollectErrorsTransRules}(\Sigma, \mathrm{P})]\!]_{Base}$, then for all x', if $x'_J = \overline{x}$ then $R(x') \notin [\![\mathrm{P}]\!]_{Base}$.

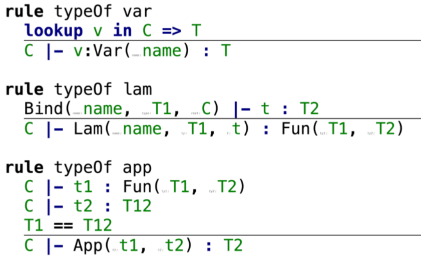
5.3 Optimizing Error Propagation

The transformation described above generates rules that propagate errors. In general, this propagation is necessary to ensure we recognize a term as ill-typed when a type error occurs in a subterm. But the propagation of errors also induces a performance overhead: If an error occurs deeply nested in a subterm, that error will be associated with the subterm and all its ancestors. Thus, when the programmer introduces or fixes a type error, the corresponding error propagation takes time.

We found that for many type systems we can eliminate error propagation. If a type system visits all nodes of the syntax tree, an explicit propagation of errors toward the root is unnecessary. Instead, we can refine well-typedness once more and require all subterms to be free of type errors: p is well-typed if and only if $p \in \mathtt{ok}$ and $(e, err) \notin \varepsilon_{\mathtt{ok}}$ for any subterm e of p and any err. With this definition it is sufficient to find the sources of errors, but it is not necessary to propagate them. We can easily adapt our transformation by removing rule *CollectErrorsTrans*-Propagate, such that error relations ε_R are only filled according to *CollectErrorsTrans*-NewError. Moreover, the resulting error relations are perfectly suited for programming editors and compilers, which can extract type errors their origin.

6 IMPLEMENTATION: A TYPE-SYSTEM DSL COMPILED TO DATALOG

We have implemented a domain-specific language (DSL) for describing textbook-like type systems. In our DSL, the programmer can declare arbitrary judgments with mixfix syntax and annotate the sorts to describe the co-functional and functional dependencies of the judgments. The actual annotations in the DSL are simpler than in the paper, but also slightly less expressive. These judgments can then be used to define type rules. The screenshot on the right shows part of a type system specification as an example of our DSL. We implemented the DSL as a metalanguage in the projectional language workbench MPS.[1] That is, our DSL can be used to define the type system of other languages defined with MPS. We developed a compiler for our DSL that generates Datalog code using the transformations described in this paper. We can target any Datalog dialect that allows Datalog rules to synthesize algebraic data (e.g., to produce function types). However, we will only obtain an incremental checker for those Datalog dialects that have an incremental solver. Specifically, we generate code conforming to the Datalog dialect of IncA [Szabó et al. 2018a, 2016], an incremental Datalog-based static analysis framework. Another Datalog dialect



```
rule typeOf var
  lookup v in C => T
  ------------------------------------------
  C |- v:Var(name) : T

rule typeOf lam
  Bind(name, T1, C) |- t : T2
  ------------------------------------------
  C |- Lam(name, T1, t) : Fun(T1, T2)

rule typeOf app
  C |- t1 : Fun(T1, T2)
  C |- t2 : T12
  T1 == T12
  ------------------------------------------
  C |- App(t1, t2) : T2
```

---

[1] https://www.jetbrains.com/mps

our approach can target is DDlog [Ryzhyk and Budiu 2019]. There are a few differences between the transformations in our implementation and the transformations described in the paper:

- In our DSL, the premises of type rules are ordered and the judgments declare input-like and output-like columns. We use this information to reason about metavariable bindings. For example, we require metavariable C to be bound before using it in a premise $C \vdash e : T$.
- In this paper, we introduce the transformations acting on Datalog code, but our compiler actually transforms type system specifications described in our DSL. Only as a last step, the compiler will lower typing rules to the Datalog code. As such, the transformations can be understood to simplify data dependencies within typing rules, which is probably useful for any incrementalization attempt, but maybe even for parallelizing type checking. Additionally, we want to emphasize that the transformations presented in this paper can be useful when implementing type checkers that do not utilize Datalog solvers. The first transformation eliminates the propagation of typing contexts and turns it into a judgment that determines the typing context to consider on demand when a variable reference is encountered. This can enable more fine-grained dependency tracking that can be utilized in other approaches to implement type checkers, including parallelizing it. In combination with the second transformation we avoid the dependencies between intermediate typing context entirely.
- Since we can reason about metavariable bindings in the implementation, slicing becomes easier. Where we used $\textbf{slice}_X(A)$ in the paper, our implementation can easily decide which atoms A are relevant.
- As usual in the type systems literature, but unlike our Datalog encodings, the conclusions of type rules in our DSL express a few syntactic constraints. Usually, these are used to dispatch the current term to the appropriate type rule. By default, our implementation of *CollectErrorsTrans* uses the constraints found in the conclusion as **IgnoreStuck**, such that no explicit declaration of **IgnoreStuck** is required.
- In addition to the transformations described in the paper, our implementation can also handle infinite relations with neither functional nor co-functional dependencies. For such relations, our implementation resolves to generating *non*-incremental Java code that can be invoked from within the Datalog rules. This is reasonable for embedding short yet intractable computations within a larger incremental computation. For example, this extension enabled us to support polymorphic types in our case studies.

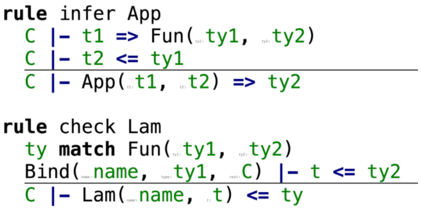
The implementation is available open source at https://gitlab.rlp.net/plmz/itypes.

7 CASE STUDIES

We conducted case studies to explore the expressivity of our DSL and of the underlying Datalog transformations. Using our DSL, we specified a range of type systems and compiled them to Datalog. In this section, we provide an overview of type system features we successfully encoded and discuss limitations.

Simple types. We encoded PCF, a simply typed lambda calculus with numeric literals, addition, if-zero, and fix. PCF extends our running example and the specification looks much the same. We also used PCF for benchmarking, which we discuss in Section 8.

Products and sums. To confirm that the DSL and Datalog transformations can handle types for compound data, we modeled product and sum types. The type rules in our DSL closely follow the rules described in *Types and Programming Languages* [Pierce 2002]. Our compiler translates the extended specification to incrementally executable Datalog code without difficulty. It is reassuring to see that our transformation rules are unchallenged by simple extensions.



```
rule infer App
  C |- t1 => Fun(ty1, ty2)
  C |- t2 <= ty1
  ------------------------------
  C |- App(t1, t2) => ty2

rule check Lam
  ty match Fun(ty1, ty2)
  Bind(name, ty1, C) |- t <= ty2
  ------------------------------
  C |- Lam(name, t) <= ty
```

Bi-directional type checking. Bi-directional type checking is a form of local type inference. The challenge of bi-directional type checking for our DSL is that there are two mutually recursive typing relations: one for checking and one for inferring types. Our transformations can handle this scenario since we never relied on the recursive structure of the typing relation, and since the underlying Datalog solver can compute mutually recursive Datalog relations. We can thus incrementalize bi-directional type systems.

Overloading. When we introduced co-functional dependencies, we argued that in an algorithmic type system all contextual information is co-functionally dependent on the syntax-tree node. This is because each syntax-tree node is visited at most once per relation (syntax-directedness), and the relevant context information must not be guessed to avoid backtracking. To explore if the type system necessarily has to be algorithmic, we modeled simple operator overloading. Specifically, we added floating-point numbers to PCF such that there are two type rules for the + operator: one for integers and one for floating-point numbers. This type system is *not* algorithmic, since we have to try out multiple type rules when encountering a + operator. Hence, the question arises whether the typing context is co-functionally dependent nonetheless, or if our transformations fail. As it turns out, overlapping type rules are not an issue for co-functional dependencies as long as all overlapping rules treat the contextual information uniformly. We believe that this usually is the case: The syntactic form governs the threading of contextual information, not the particular type rule applied.

Universal types. We extended PCF with universal types in the style of System F. The main challenge for incrementality and our transformations is the substitution function on types, that the type system uses to instantiate universal types. As a relation, type substitution takes the form `tsubst` $: T \times X \times T \times T$ for types T and type variables X. This relation is infinite and there are no co-functional dependencies. Therefore, our transformations cannot make this relation compatible with bottom-up evaluation and thus not incremental. In such cases, our implementation falls back to generating non-incremental Java code that is being invoked from within the incremental Datalog code. This is acceptable when only a small portion of the overall computation becomes non-incremental. For universal types, only type substitution is non-incremental, while tree traversal, variable lookups, type propagation, etc. are fully incremental. However, the size of types usually does not grow proportionally with the size of the AST, such that non-incremental type substitution does not endanger the incremental performance. Technically, our incremental Datalog solver selectively reruns non-incremental operations when any of their arguments changes.

Iso-recursive types. We can also support iso-recursive types as introduced in *Types and Programming Languages* [Pierce 2002]. The typing rules for the fold and unfold expressions depend on type substitution just like universal types. The extension of iso-recursive types does not violate the property of being an algorithmic type system. With this extension and the already supported sum, product and universal types, we can support an expressive functional programming languages type system.

Limitations. We are aware of a few limitations that we want to disclose. First, our DSL currently does not provide support for handling lists, which makes it difficult to encode type system features such as records, variants, or functions with multiple parameters. This is a DSL limitation, not a limitation of our approach of generating incremental Datalog programs. Second, since type substitution is difficult to incrementalize (see universal types), unification is difficult to incrementalize. Therefore, it is not clear if type systems with Hindley-Milner type inference can be supported by our approach. Third, as of right now our approach does not support dependent types. In order to support dependent types our approach needs to be extended such that it can translate operational semantics to Datalog. As a first experiment we were able to encode an interpreter for PCF in the Datalog dialect IncA which enables us to incrementalize operational semantics. Hence, it is possible to describe an type checker supporting dependent types in Datalog. Lastly, we investigated if our approach can support languages with nominal subtyping. Like type substitution, subtyping is an infinite relation without co-functional dependencies, and we must resolve to generating non-incremental Java code. However, nominal subtyping is not self-contained and requires access to the class table of the program in order to decide $C <: D$. This induces additional constraints on when to rerun a subtype check, which we cannot currently trace.

8 PERFORMANCE EVALUATION

We present a preliminary performance evaluation of our approach using a type checker for PCF and synthesized subject programs. Our goal is to assess the incremental performance of our approach, and to examine that impact of our transformation steps on the performance. We compare to a non-incremental recursive descent type checker written in Java.

We synthesize two PCF programs *Star* and *Chain* that have intricate dependencies and challenge our incremental approach. *Star* consists of n functions all calling f_0. *Chain* consists of n functions each calling f_{n-1}. These programs allow us to introduce changes with global effect on type checking.

Star:
$\text{let } f_0 = \lambda x{:}\mathsf{Nat}.\, 1 + x \text{ in}$
$\quad \text{let } f_1 = \lambda x{:}\mathsf{Nat}.\, 1 + f_0(x),$
$\qquad \ldots,$
$\qquad f_n = \lambda x{:}\mathsf{Nat}.\, 1 + f_0(x) \text{ in}$
$\qquad\quad 1 + f_0(1)$

Chain:
$\text{let } f_0 = \lambda x{:}\mathsf{Nat}.\, 1 + x \text{ in}$
$\quad \text{let } f_1 = \lambda x{:}\mathsf{Nat}.\, 1 + f_0(x) \text{ in}$
$\qquad \ldots$
$\qquad \text{let } f_n = \lambda x{:}\mathsf{Nat}.\, 1 + f_{n-1}(x) \text{ in}$
$\qquad\quad 1 + f_n(1)$

We generate small IDE-style program changes (as opposed to larger commit-style changes) for our evaluation. Our changes are local and only affect a single subterm. We change the program by deleting and inserting tuples into the sets of tuples that describe the program, but only those that are directly affected by the change. The incremental Datalog solver will make sure to propagate those deletions and insertions as to update all derived tuples. To stress-test our approach, we always apply changes to f_0, which all other functions (transitively) depend on. We consider the following 6 kind of changes and their inverse undo changes:

- *Num*: Increment the value of a numeric literal by 1.
- *Ref*: Change a variable reference to an unbound name.
- *Param*: Change the parameter name of a lambda abstraction.
- *Anno*: Change the type annotation of a lambda abstraction.
- *Lambda*: Insert a lambda abstraction in the body of an existing lambda abstraction.
- *AddApp*: Change an addition to an application while retaining the original operands.

Note that, except for *Num*, all of the above changes will result in an ill-typed program. We believe this realistically reflects programming sessions, where a developer changes one piece at a time.

We consider 4 type checker implementations:

- B: baseline type checker, non-incremental, written as a recursive Java function.
- T1: incremental checker, only using our first transformation *CoFunTrans*.
- T1+T2: incremental checker, additionally using our second transformation *CtxFusionTrans*.
- T1+T2+T3: incremental checker, additionally using our third transformation *CollectErrorsTrans*.

For the measurements, we synthesize subject programs *Star* and *Chain* with $n = 200$. We apply each change and its undo 40 times after warmup. We measure the initial analysis time and the time

Program	Checker	Initial	Num	Ref	Param	Anno	Lambda	AddApp
Star	B	6.24						
	T1	260.48	0.02±0.00	35.11±1.23	34.42±1.11	36.86±1.15	61.16±0.96	34.92±1.36
	+T2	314.74	0.07±0.05	32.56±1.46	34.37±1.46	31.83±1.22	53.14±0.66	30.91±1.36
	+T3	320.91	0.06±0.00	0.17±0.02	0.10±0.00	43.91±0.57	42.96±0.57	20.98±0.78
Chain	B	49.31						
	T1	176.00	0.06±0.02	85.61±8.58	127.26±4.25	126.99±4.74	44.44±2.00	47.32±2.32
	+T2	1016.76	0.02±0.00	82.45±3.80	79.02±3.45	87.60±4.69	87.27±4.42	82.00±3.89
	+T3	1040.44	0.02±0.00	0.08±0.00	0.096±0.00	1.46±0.07	1.71±0.08	1.16±0.07

Fig. 4. Summary of the measurement results. All values are in milliseconds. For average update times, we also show the 95% confidence interval. B stands for the non-incremental baseline type checker. T1 is short for *CoFunTrans*, T2 is *CtxFusionTrans*, and T3 is *CollectErrorsTrans*. Our DSL yields the (T1+T2)+T3 running times.

it takes to process a change. We performed our benchmarks on a machine with an Intel Core i7 at 2.7 GHz with 16 GB of RAM, running 64-bit OSX 10.15.4, Java 1.8.0_222, and MPS version 2019.1.6.

Results. Figure 4 shows a summary of our measurement results. First, let us discuss the performance of the final transformation stage T1+T2+T3 that our DSL uses and compare it to the non-incremental baseline type checker. We observe that the incremental update times are really fast; they are at most several tens of milliseconds, which is exactly what we expect from a type checker running in an IDE. The initialization time is at most a second, which we consider acceptable, as this is a one-time cost. The run time of the baseline analysis is also fast. This is not surprising because our subject programs are small. However, incrementalization brings significant performance gains most of the time compared to the baseline version. For example, for *Num*, *Ref*, and *Param* changes we see multiple orders of magnitude speedups. We also see slowdowns in certain cases: For the *Star* program the *Anno*, *Lambda*, and *AddApp* changes induce an order of magnitude slowdown. This is due to the global effect these changes have on the program, which our incremental analysis has to retrace. But even in these cases, the incremental running times are still much faster than the initial run of our analysis. In future work, we will try to speed up the initial analysis run, which should improve the performance for changes with global effect.

Let us now examine the effect of our transformations on the performance. For *CtxFusionTrans* (T2), the *Chain* program is interesting because it requires a long threading of typing contexts. When we change the type of f_0 (changes *Param* and *Anno*), all threaded contexts become invalid. However, for changes *Lambda* and *AddApp* we can observe a negative effect of *CtxFusionTrans*. This is because those changes eliminate the binding of f_0 altogether, since its definition becomes ill-typed. Transformation *CollectErrorsTrans* (T3) recovers these losses. Indeed, *CollectErrorsTrans* (T3) induces a significant speedup most of the time. This is because error collection makes the entire `typeOf` relation more resilient to changes. That is, the type checker can endure ill-typed terms and reuse the tuples in the relation more frequently. As it turns out, this separation of type inference and error collection is key to fast incremental type checking.

Given that incrementalization comes with extensive caching, we also benchmarked the memory overhead of our type checkers. We found that on average the memory overhead is around 10 MB, which is a negligible value compared to the 2 GB memory consumption of the IDE itself.

To summarize, we find that the incremental performance of our type checker is suitable for applications in IDEs. We often achieve order-of-magnitude speedups compared to the non-incremental baseline analysis. We pay the price for this with occasional slowdowns in update times and longer initialization time. The memory overhead of our approach is negligible for the synthesized programs.

9 RELATED WORK

IncA [Szabó et al. 2016] is an incremental static analysis framework based on Datalog. We use IncA in our work as the incremental evaluation engine for the Datalog code we generate. Specifically, we show how to systematically construct type checkers that can then be automatically incrementalized by IncA. IncA has been shown previously to deliver fast incremental updates for a range of program analyses: FindBugs-style linting, control-flow analysis [Szabó et al. 2016], data-flow analyses [Szabó et al. 2018a], and overload resolution [Szabó et al. 2018b]. This paper is the first to generate IncA code from high-level specifications.

Typol [Despeyroux 1984] translates inference rules to Prolog. In contrast to Datalog, Prolog is a Turing-complete language and supports infinite relations that are explored on-demand through top-down evaluation. The same holds for other works such as [Farka et al. 2018; Franceschini et al. 2016] which compile to first-order Horn clauses where infinte relations are allowed. Thus, we face the more difficult challenges of translating inference rules into a style that permits the bottom-up evaluation of logic programs. Attali et al. [1992] implemented an incremental evaluator for Typol programs. However, for fast incremental update times, the context is not allowed to change. If the context changes in any way, type checking has to be started from scratch for the affected expressions. In contrast, we derive a Datalog program that is resilient to such changes.

Wachsmuth et al. [2013] propose a task engine for incremental name and type analysis. Tasks tend to be small and inter-dependent, encoding fine-grained dependencies. When a file changes, they (re)generate tasks for the entire file. Task evaluation relies on a cache of previous task results, only recomputing tasks that are new. If a change affects a task, its cache entry is invalidated and the task reevaluated. The task engine then triggers the reevaluation of all transitively dependent tasks. In contrast to this specialized approach, we rely on a generic incremental compilation target, namely Datalog. The transformations we presented in this paper enable us to handle type systems based on standard typing rules, whereas the task engine requires language-specific rules for task generation.

Erdweg et al. [2015] introduce a co-contextual formulation of type checking. Similar to our approach, co-contextual type checking eliminates context propagation. However, while we synthesize a find relation to lookup bindings as needed, co-contextual type checkers propagate lookup constraints when encountering a variable. This makes co-contextual type checkers compositional, allowing subderivations to be reused even when context information changes. The caveat of co-contextual type checking is that incremental performance heavily relies on constraints being locally solved in the subderivations, which often is not the case [Kuci et al. 2017]. In our solution, we use Datalog's dependency tracking instead of trying to fit dependencies into a compositional structure.

The work on incremental type checking for the programming language B [Meertens 1983] decorates the syntax tree with the type requirements known for a specific node. When changing a node in the syntax tree, the decorated node is deleted which is followed by inserting the newly decorated node while reusing the decorated children of the deleted node. This technique only works because B does not support type declarations for variables but infers the type by discovering type requirements based on the usage of the variable. Hence, the type system of B does not require top-down context propagation, which we support by utilizing *co-functional dependencies*. As our case studies indicate, our approach is applicable to many type systems.

Busi et al. [2019] propose to incrementalize type checking by deriving type rules that utilize memoization. This allows the reuse of parts of the typing derivation when code changes occur. However, as soon as any part of the context or the expression is changed, the entire subderivation has to be reconstructed. Our approach uses much more fine-grained dependency tracking. In particular, our second transformation enables us to track individual bindings rather than entire contexts, which our evaluation confirmed to be essential for incremental type checking.

Datafun [Arntzenius and Krishnaswami 2020] is a higher-order functional language that incorporates Datalog's semi-naive bottom-up evaluation. While Datafun does not aim for incrementality, it would be interesting to see if our transformations can expand the expressivity of Datafun. Specifically, it would be interesting to demonstrate that the bottom-up computable Datalog code we generate indeed is admitted by Datafun's type system. Their type system enforces monotonicity constraints that our Datalog solver relies on, too.

The transformation we presented in Section 3 has a strong resemblance with magic set transformations [Beeri and Ramakrishnan 1991], which are well-known in the Datalog community. Like our transformation, a magic set transformation rewrites a Datalog program to eliminate the derivation of irrelevant (unquerried) tuples. Traditionally, magic set transformations are used as an optimization that may filter some or all irrelevant tuples, and usually the original Datalog program is already computable (has finite relations). In contrast, we start with an incomputable Datalog program (infinite relations). Therefore, we developed a specialized transformation that exploits the new concept of co-functional dependencies. Our specialized transformation allows us to guarantee all irrelevant tuples are eliminated and that the typing relation becomes finite. Additionally, our transformation can exploit co-functional dependencies to avoid the generation of additional auxiliary relations that traditional magic set transformations would require.

Deforestation [Wadler 1990] is a technique to avoid intermediate immutable data structures that are produced and immediately consumed. The *context fusion* transformation of Section 4 follows the same idea. Instead of constructing intermediate maps (e.g. typing contexts) and letting the built-in `lookup` relation consume them, we directly perform lookup on the data that functionally determines the map that is passed to `lookup`. The consumer is fixed (`lookup`) in our approach, but every relation that functionally determines a map is viable as a producer. Our technique is applicable to recursive Datalog programs while deforestation is an optimization technique for functional programs.

10 CONCLUSION

We proposed a novel approach to systematically deriving incremental type checkers based on textbook-style type rules. Our solution is divided into three different transformations. The first transformation utilizes *co-functional dependencies* to translate type rules to Datalog such that bottom-up evaluation succeeds. The second transformation eliminate dependencies on compound data such as typing contexts to achieve more efficient incremental performance. And the third transformation separates the error collection from the type rules, which is interesting even outside of this work. While our transformations primarily tackle issues with incremental name bindings and type errors, we showcased that our transformations effectively support a wide range of type system features such as sum and product types, overloading, universal types, iso-recursive types, and bi-directional type checking. Further, we performed a preliminary performance evaluation to demonstrate that the derived type checkers indeed achieve fast incremental update times.

ACKNOWLEDGEMENTS

We thank Sven Keidel, Gábor Bergmann, and the anonymous reviewers for their comments.

REFERENCES

Mario Alvarez-Picallo, Alex Eyers-Taylor, Michael Peyton Jones, and C.-H. Luke Ong. 2019. Fixing Incremental Computation - Derivatives of Fixpoints, and the Recursive Semantics of Datalog. In *Programming Languages and Systems - 28th European*

Symposium on Programming, ESOP 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings (Lecture Notes in Computer Science), Luís Caires (Ed.), Vol. 11423. Springer, 525–552. https://doi.org/10.1007/978-3-030-17184-1_19

Michael Arntzenius and Neel Krishnaswami. 2020. Seminaïve evaluation for a higher-order functional language. *Proc. ACM Program. Lang.* 4, POPL (2020), 22:1–22:28. https://doi.org/10.1145/3371090

Isabelle Attali, Jacques Chazarain, and Serge Gilette. 1992. Incremental Evaluation of Natural Semantics Specification. In *Programming Language Implementation and Logic Programming, 4th International Symposium, PLILP'92, Leuven, Belgium, August 26-28, 1992, Proceedings (Lecture Notes in Computer Science)*, Maurice Bruynooghe and Martin Wirsing (Eds.), Vol. 631. Springer, 87–99. https://doi.org/10.1007/3-540-55844-6_129

Catriel Beeri and Raghu Ramakrishnan. 1991. On the Power of Magic. *J. Log. Program.* 10, 3&4 (1991), 255–299. https://doi.org/10.1016/0743-1066(91)90038-Q

Matteo Busi, Pierpaolo Degano, and Letterio Galletta. 2019. Using Standard Typing Algorithms Incrementally. In *NASA Formal Methods - 11th International Symposium, NFM 2019, Houston, TX, USA, May 7-9, 2019, Proceedings (Lecture Notes in Computer Science)*, Julia M. Badger and Kristin Yvonne Rozier (Eds.), Vol. 11460. Springer, 106–122. https://doi.org/10.1007/978-3-030-20652-9_7

Thierry Despeyroux. 1984. Executable Specification of Static Semantics. In *Semantics of Data Types, International Symposium, Sophia-Antipolis, France, June 27-29, 1984, Proceedings (Lecture Notes in Computer Science)*, Gilles Kahn, David B. MacQueen, and Gordon D. Plotkin (Eds.), Vol. 173. Springer, 215–233. https://doi.org/10.1007/3-540-13346-1_11

Sebastian Erdweg, Oliver Bracevac, Edlira Kuci, Matthias Krebs, and Mira Mezini. 2015. A co-contextual formulation of type rules and its application to incremental type checking. In *Proceedings of the 2015 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2015, part of SPLASH 2015, Pittsburgh, PA, USA, October 25-30, 2015*, Jonathan Aldrich and Patrick Eugster (Eds.). ACM, 880–897. https://doi.org/10.1145/2814270.2814277

Frantisek Farka, Ekaterina Komendantskaya, and Kevin Hammond. 2018. Proof-relevant Horn Clauses for Dependent Type Inference and Term Synthesis. *Theory Pract. Log. Program.* 18, 3-4 (2018), 484–501. https://doi.org/10.1017/S1471068418000212

Luca Franceschini, Davide Ancona, and Ekaterina Komendantskaya. 2016. Structural Resolution for Abstract Compilation of Object-Oriented Languages. In *Proceedings of the First Workshop on Coalgebra, Horn Clause Logic Programming and Types, CoALP-Ty 2016, Edinburgh, UK, 28-29 November 2016 (EPTCS)*, Ekaterina Komendantskaya and John Power (Eds.), Vol. 258. 19–35. https://doi.org/10.4204/EPTCS.258.2

Sylvia Grewe, Sebastian Erdweg, Pascal Wittmann, and Mira Mezini. 2015. Type systems for the masses: deriving soundness proofs and efficient checkers. In *2015 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, Onward! 2015, Pittsburgh, PA, USA, October 25-30, 2015*, Gail C. Murphy and Guy L. Steele Jr. (Eds.). ACM, 137–150. https://doi.org/10.1145/2814228.2814239

Robert Harper. 2016. *Practical Foundations for Programming Languages (2nd. Ed.).* Cambridge University Press. https://www.cs.cmu.edu/%7Erwh/pfpl/index.html

Edlira Kuci, Sebastian Erdweg, Oliver Bracevac, Andi Bejleri, and Mira Mezini. 2017. A Co-contextual Type Checker for Featherweight Java. In *31st European Conference on Object-Oriented Programming, ECOOP 2017, June 19-23, 2017, Barcelona, Spain (LIPIcs)*, Peter Müller (Ed.), Vol. 74. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 18:1–18:26. https://doi.org/10.4230/LIPIcs.ECOOP.2017.18

Lambert G. L. T. Meertens. 1983. Incremental Polymorphic Type Checking in B. In *Conference Record of the Tenth Annual ACM Symposium on Principles of Programming Languages, Austin, Texas, USA, January 1983*, John R. Wright, Larry Landweber, Alan J. Demers, and Tim Teitelbaum (Eds.). ACM Press, 265–275. https://doi.org/10.1145/567067.567092

Benjamin C. Pierce. 2002. *Types and programming languages.* MIT Press.

Raghu Ramakrishnan, François Bancilhon, and Abraham Silberschatz. 1987. Safety of Recursive Horn Clauses With Infinite Relations. In *Proceedings of the Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, March 23-25, 1987, San Diego, California, USA*, Moshe Y. Vardi (Ed.). ACM, 328–339. https://doi.org/10.1145/28659.28694

Leonid Ryzhyk and Mihai Budiu. 2019. Differential Datalog. In *Datalog 2.0 2019 - 3rd International Workshop on the Resurgence of Datalog in Academia and Industry co-located with the 15th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2019) at the Philadelphia Logic Week 2019, Philadelphia, PA (USA), June 4-5, 2019 (CEUR Workshop Proceedings)*, Mario Alviano and Andreas Pieris (Eds.), Vol. 2368. CEUR-WS.org, 56–67. http://ceur-ws.org/Vol-2368/paper6.pdf

Yannis Smaragdakis and Martin Bravenboer. 2010. Using Datalog for Fast and Easy Program Analysis. In *Datalog Reloaded - First International Workshop, Datalog 2010, Oxford, UK, March 16-19, 2010. Revised Selected Papers (Lecture Notes in Computer Science)*, Oege de Moor, Georg Gottlob, Tim Furche, and Andrew Jon Sellers (Eds.), Vol. 6702. Springer, 245–251. https://doi.org/10.1007/978-3-642-24206-9_14

Tamás Szabó, Gábor Bergmann, Sebastian Erdweg, and Markus Voelter. 2018a. Incrementalizing lattice-based program analyses in Datalog. *PACMPL* 2, OOPSLA (2018), 139:1–139:29. https://doi.org/10.1145/3276509

Tamás Szabó, Sebastian Erdweg, and Markus Voelter. 2016. IncA: a DSL for the definition of incremental program analyses. In *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering, ASE 2016, Singapore, September 3-7, 2016*, David Lo, Sven Apel, and Sarfraz Khurshid (Eds.). ACM, 320–331. https://doi.org/10.1145/2970276.2970298

Tamás Szabó, Edlira Kuci, Matthijs Bijman, Mira Mezini, and Sebastian Erdweg. 2018b. Incremental overload resolution in object-oriented programming languages. In *Companion Proceedings for the ISSTA/ECOOP 2018 Workshops, ISSTA 2018, Amsterdam, Netherlands, July 16-21, 2018*, Julian Dolby, William G. J. Halfond, and Ashish Mishra (Eds.). ACM, 27–33. https://doi.org/10.1145/3236454.3236485

Zoltán Ujhelyi, Gábor Bergmann, Ábel Hegedüs, Ákos Horváth, Benedek Izsó, István Ráth, Zoltán Szatmári, and Dániel Varró. 2015. EMF-IncQuery: An integrated development environment for live model queries. *Sci. Comput. Program.* 98 (2015), 80–99. https://doi.org/10.1016/j.scico.2014.01.004

Guido Wachsmuth, Gabriël D. P. Konat, Vlad A. Vergu, Danny M. Groenewegen, and Eelco Visser. 2013. A Language Independent Task Engine for Incremental Name and Type Analysis. In *Software Language Engineering - 6th International Conference, SLE 2013, Indianapolis, IN, USA, October 26-28, 2013. Proceedings (Lecture Notes in Computer Science)*, Martin Erwig, Richard F. Paige, and Eric Van Wyk (Eds.), Vol. 8225. Springer, 260–280. https://doi.org/10.1007/978-3-319-02654-1_15

Philip Wadler. 1990. Deforestation: Transforming Programs to Eliminate Trees. *Theor. Comput. Sci.* 73, 2 (1990), 231–248. https://doi.org/10.1016/0304-3975(90)90147-A

Adrienne Watt. 2018. *Database design.*