Mapping Polymorphism - Proofs

Ryan Wisnesky
Mauricio A. Hernandez
and
Lucian Popa

TR-10-09

Computer Science Group
Harvard University
Cambridge, Massachusetts

Mapping Polymorphism - Proofs

Ryan Wisnesky
Harvard University
ryan@cs.harvard.edu

Mauricio A. Hernández, Lucian Popa
IBM Almaden Research Center
{mauricio, lucian}@almaden.ibm.com

Abstract

This technical report contains the proofs for the paper *Mapping Polymorphism*, by Ryan Wisnesky, Mauricio A. Hernández, and Lucian Popa, which appears in the proceedings of the 13th International Conference on Database Theory, held March 22–25, 2010, in Lausanne, Switzerland (ICDT '10).

Theorem 1 - Satisfiability

Proof. We follow the general technique of constructing canonical databases (or canonical models). While there is some analogy with satisfiability of conjunctive queries [1], our proof is more involved, since our core mapping language is more complex, involves nested relational data, and uses both universal and existential quantification.

We construct a canonical model by initializing first an environment where all top-level sets are empty. More precisely, for each schema root S we create an instance I as follows. If S is a record type, then I must contain all the labels given by the record type. For each label l, we create an instance of the corresponding type t. If t is another record type, we recurse. If t is a `SETRCD` or `SETCHC` type, the instance we create is an empty set. If t is `ATOMIC` a, we create a fresh new "symbol" (or labeled null) α of type a. After this initialization, all such symbols have unique occurrences.

This initial environment is then enriched and modified based on the typing derivation of M as follows. Whenever we encounter an existentially quantified subexpression (in the typing derivation) we construct a new instance, standing for the existentially quantified variable, and insert it in the corresponding set in the environment. The new instance (either a record or a choice) is minimally constructed in the sense that we populate its required structure while leaving empty any of its nested sets. The construction is the same as in the initialization phase. Atomic type fields (whenever they appear) are initialized with fresh new symbols or nulls.

Whenever we encounter equalities, implications or universally quantified subexpressions as part of the typing derivation of M, we alter the existing environment via a process similar to the chase with tgds and egds that constructs a canonical universal solution [2]. Equalities are used to identify the symbolic values of atomic fields. Since there are no equalities between constants in the language, and the instances are entirely symbolic, this process will never fail. For a universally quantified subexpression, we need to "evaluate" the quantifier in all possible ways on the current environment and enforce the body of the subexpression. This enforcement may require enforcing new equalities or, recursively, creating new instances for existentially quantified subexpressions.

We illustrate with a concrete example. Consider the following core mapping expression:

$$M = \underline{\mathsf{exists}}\ x\ \underline{\mathsf{in}}\ R.\ \underline{\mathsf{exists}}\ y\ \underline{\mathsf{in}}\ S.\ x.A = y.A\ \wedge\ (\underline{\mathsf{for}}\ x'\ \underline{\mathsf{in}}\ R.\ x'.A = y.A \Rightarrow x'.B = y.B)$$

The mapping M type-checks in a context $\Gamma = (R, \texttt{SETRCD}(\!|\texttt{A} : \texttt{ATOMIC a}, \texttt{B} : \texttt{ATOMIC b}|\!)); (\texttt{S}, \texttt{SETRCD}(\!|\texttt{A} : \texttt{ATOMIC a}, \texttt{B} : \texttt{ATOMIC b}|\!))$. We construct an environment $\Delta \in [\![\Gamma]\!]$ as follows. At the beginning, $\Delta = (R, \emptyset); (S, \emptyset)$. After processing the first two existential quantifiers, we have that R contains a record $\{A : \alpha, B : \beta\}$ and S contains a record $\{A : \alpha', B : \beta'\}$, where α and α' are fresh new nulls of type `ATOMIC` a, and β and β' are fresh new nulls of type `ATOMIC` b. After processing the equality $x.A = y.A$, the two nulls α and α' are equated and we replace one of them with the other, say α' with α. When we encounter the universally quantified expression, we evaluate the binder in the existing environment. In this case, x' binds to the single record of R. We then enforce the inner implication. Since the equality $x'.A = y.A$ is satisfied (both expressions evaluate to α), we must enforce the equality $x'.B = y.B$. As a result, we obtain that β and β' must be equal and we replace β' with β. The resulting environment will have both R and S consist of the single record $\{A : \alpha, B : \beta\}$.

At the end of such construction, we will replace all symbols with distinct constants from the domain. The resulting environment is a valid environment that satisfies M. □

Proposition 1 - Principal Typings are Unique

Proof. Write $=_\alpha$ to indicate alpha equivalence (equivalence up to one-to-one renaming of type variables). First, we need a lemma:

$$\forall\ t_1\ t_2\ \phi_{12}\ \phi_{21},\ \phi_{12}t_1 = t_2 \land \phi_{21}t_2 = t_1 \to t_1 =_\alpha t_2$$

Proceed by mutual induction over schema and rows. The predicates for mutual induction are the same for the row and schema cases. Schema are the easy case:

- When t_1 is some variable v, since $\phi_{21}t_2 = v$ it must be the case that t_2 is some other variable u, and so $t_1 =_\alpha t_2$.
- When $t_1 = C\ r$, we know that

  $$\phi_{12}\big(C\ r\big) = t_2 \land \phi_{21}t_2 = C\ r$$

  We know that $t_2 = C\ r'$ for some r', since the variable case is impossible, and we know that

  $$\phi_{12}r = r' \land \phi_{21}r' = r$$

  As such we can apply the inductive hypothesis over rows to obtain $r =_\alpha r'$, from which it follows that $t_1 =_\alpha t_2$.

For the row case,

- When r_1 is some variable v, since $\phi_{21}r_2 = v$ it must be the case that r_2 is some other variable u, and so $r_1 =_\alpha r_2$.
- When r_1 is $\llparenthesis\rrparenthesis$, since $\phi_{12}\llparenthesis\rrparenthesis = r_2$ it must be the case that r_2 is $\llparenthesis\rrparenthesis$, and so $r_1 =_\alpha r_2$.
- When r_1 is $\llparenthesis l : \tau, r\rrparenthesis$, we know that

  $$\phi_{12}\big(\llparenthesis l : \tau, r\rrparenthesis\big) = r_2 \land \phi_{21}r_2 = \llparenthesis l : \tau, r\rrparenthesis$$

  We know that r_2 can't be the empty row or a type variable because of the left conjunct, so r_2 must be $\llparenthesis l : \tau', r'\rrparenthesis$. Thus we know that

  $$(A) \quad \phi_{12}\tau = \tau' \land \phi_{12}r = r' \land \phi_{21}\tau' = \tau \land \phi_{21}r' = r$$

  Our inductive hypotheses thus yield that

  $$\tau =_\alpha \tau' \land r =_\alpha r'$$

  Finally, to show that the above implies $t_1 =_\alpha t_2$, we will merge the two α-equivalence substitutions above (call them α and α') to obtain a new α-equivalence substitution. We have

  $$\alpha\tau = \tau' \land \alpha\tau' = \tau \land \alpha' r = r' \land r = \alpha' r$$

  We just need to show that α and α' agree on the type variables occurring in common in τ, τ', r, r'. From above and the leftmost two conjuncts of (A) we know that for any such v, $\alpha v = \phi_{12}v$ and $\alpha' v = \phi_{12}v$, whence $\alpha = \alpha'$.

It is easy to apply the lemma to get uniqueness of principal typings. Suppose we have two principal typings $\Gamma_1 = (v : t_1, \ldots)$ and $\Gamma_2 = (v : t_2, \ldots)$. We know from the principal typings property that there exists substitions ϕ_{12} and ϕ_{21} such that $\phi_{12}\Gamma_1 = \Gamma_2$ and $\phi_{21}\Gamma_2 = \Gamma_1$. Because the variables in their domains are unique, we can apply the above lemma to obtain that

$$\texttt{RCD}\ \llparenthesis v : t_1, \ldots\rrparenthesis =_\alpha \texttt{RCD}\ \llparenthesis v : t_2, \ldots\rrparenthesis$$

It follows that $\Gamma_1 =_\alpha \Gamma_2$.

□

Proposition 2 - Schema Unification produces MGUs

Proof. Our schema unification rules are simply those of [3] specialized to NR schema. Hence this is a straightforward consequence of this property in that system. □

Theorem 2 - Soundness

We start by establishing the soundness of path inference over path checking. We will need:

Lemma 1 (Path checking respects substitution)**.** $\forall \Gamma\ p\ t\ \phi,\ \Gamma \vdash p :: t \rightarrow \phi\Gamma \vdash p :: \phi t.$

Proof. Introduce Γ and proceed by induction on p.

- In the case where p is a variable, $(v, t) \in \Gamma$ and so $(v, \phi t) \in \phi\Gamma$ and the result holds by VAR.
- In the case where we have a path $p.l$, the inductive hypothesis is that

$$\forall t\ \phi,\ \Gamma \vdash p :: t \rightarrow \phi\Gamma \vdash p :: \phi t$$

  Introduce Γ and t and ϕ and assume that $\Gamma \vdash p.l :: t$. By inversion there is some r such that $\Gamma \vdash p :: \texttt{RCD}\ (\!|l : t, r|\!)$ Our goal is that $\phi\Gamma \vdash p.l :: \phi t$. Apply the inductive hypothesis to get $\phi\Gamma \vdash p :: \phi\texttt{RCD}\ (\!|l : t, r|\!)$. The result then follows by the RCD-ELIM rule.

□

Also recall the definition of a unifier: if $a \overset{\phi}{\sim} b$, then $\phi a = \phi b$. Carrying on then, we have:

Lemma 2 (Soundness of Path Inference)**.** $\forall\ \Gamma\ p\ \varphi\ \tau,\ \varphi\Gamma \Vdash p :: \tau \rightarrow \varphi\Gamma \vdash p :: \tau.$

Proof. Introduce Γ and proceed by induction on p.

The base case is that p is a variable. In this case, the checking and inference rules are identical. (The substitution returned from the inference algorithm is the identity function.)

The inductive step is that p is a projection. The inductive hypothesis is $\forall \varphi\tau, \varphi\Gamma \Vdash p :: \tau \rightarrow \varphi\Gamma \vdash p :: \tau$. We must show that $\forall \varphi\tau, \varphi\Gamma \Vdash p.l :: \tau \rightarrow \varphi\Gamma \vdash p.l :: \tau$. Introduce φ and τ and assume that $\varphi\Gamma \Vdash p.l :: \tau$. By inversion, we know that there exists $\phi, \psi, t, \sigma, \rho$ such that $\varphi = \psi \circ \phi$ and $\tau = \psi\sigma$ and $\phi\Gamma \Vdash p :: t$ and $\texttt{RCD}\ (\!|l :: \sigma, \rho|\!) \overset{\psi}{\sim} t$ with σ, ρ fresh. Substituting for τ and φ gives us a goal of $\psi\phi\Gamma \vdash p.l :: \psi\sigma$.

We now apply the RCD-ELIM rule, setting its Γ to be $\psi\phi\Gamma$, and its t and r to be $\psi\sigma$ and $\psi\rho$, respectively. This yields a new goal of $\psi\phi\Gamma \vdash p :: \texttt{RCD}\ (\!|l : \psi\sigma, \psi\rho|\!)$. Using our inductive hypothesis with ϕ and t and $\phi\Gamma \Vdash p :: t$ gives us that $\phi\Gamma \vdash p :: t$. We can then apply ψ to both sides (by Lemma 1) to get that $\psi\phi\Gamma \vdash p :: \psi t$. By the mgu (most general unifier) property we know that $\psi\ \texttt{RCD}\ (\!|l : \sigma, \rho|\!) = \psi t$, and we're done. □

Similarly to the case for paths, we need that

Lemma 3 (Type checking respects substitution)**.** $\forall m\ \Gamma\ \phi, \Gamma \vdash m \rightarrow \phi\Gamma \vdash m.$

Proof. By induction on m.

- The $\top$ case is trivial, as any context will work with the TRUE rule.
- For the $m_1 \oplus m_2$ case, we have two inductive hypotheses

$$\forall \Gamma\ \phi,\ \Gamma \vdash m_1 \rightarrow \phi\Gamma \vdash m_1$$

  and

$$\forall \Gamma\ \phi,\ \Gamma \vdash m_2 \rightarrow \phi\Gamma \vdash m_2$$

  Introduce Γ and ϕ and assume that $\Gamma \vdash m_1 \oplus m_2$. We want to prove that $\phi\Gamma \vdash m_1 \oplus m_2$. By inversion, we have that $\Gamma \vdash m_1$ and $\Gamma \vdash m_2$. Apply these with the inductive hypothesis gives $\phi\Gamma \vdash m_1$ and $\phi\Gamma \vdash m_2$, and the result follows from the WF-ANDIMPL rule.

- For the $p_1 \ = \ p_2$ case, we have no inductive hypothesis. Introduce Γ and ϕ and assume that $\Gamma \vdash p_1 \ = \ p_2$. By inversion, there is some a such that $\Gamma \vdash p_1 :: \mathtt{ATOMIC}\ a$ and $\Gamma \vdash p_2 :: \mathtt{ATOMIC}\ a$. Because path checking respects substitution, we have that $\phi\Gamma \vdash p_1 :: \mathtt{ATOMIC}\ \phi a$ and $\phi\Gamma \vdash p_2 :: \mathtt{ATOMIC}\ \phi a$. We can then apply the WF-EQ rule using ϕa and $\phi\Gamma$ to obtain our goal that $\phi\Gamma \vdash p_1 \ = \ p_2$.

- For the $\diamond\ v\ \underline{\mathsf{in}}\ p.\ m$ case, the inductive hypothesis is that

  $$\forall \Gamma\ \phi,\ \Gamma \vdash m \to \phi\Gamma \vdash m$$

  Introduce Γ and ϕ and assume that $\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. We want to prove that $\phi\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. By inversion we know that there is some r such that $\Gamma \vdash p :: \mathtt{SETRCD}\ r$ and $(v, \mathtt{RCD}\ r); \Gamma \vdash m$. Apply the inductive hypothesis to get $\phi\ (v, \mathtt{RCD}\ r); \Gamma \vdash m$. By the SETRCD-ELIM rule, we thus just need to obtain $\phi\Gamma \vdash p :: \phi\mathtt{SETRCD}\ r$, which follows from the soundness of path checking.

- For the $\diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$ case, the inductive hypothesis is that

  $$\forall \Gamma\ \phi,\ \Gamma \vdash m \to \phi\Gamma \vdash m$$

  Introduce Γ and ϕ and assume that $\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. We want to prove that $\phi\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. By inversion we know that there is some r and t such that $\Gamma \vdash p :: \mathtt{SETCHC}\ (\!| l : t, r |\!)$ and $(v, t); \Gamma \vdash m$. Apply the inductive hypothesis to get $\phi\ (v, t); \Gamma \vdash m$. By the SETRCD-ELIM rule, we thus just need to obtain $\phi\Gamma \vdash p :: \phi\mathtt{SETCHC}\ (\!| l : t, r |\!)$, which follows from the soundness of path checking.

□

The main result:

Lemma 4 (Soundness of Type Inference)**.** *$\forall\ m\ \Gamma\varphi,\ \varphi\Gamma \Vdash m \to \varphi\Gamma \vdash m.$*

Proof. The proof is by induction on m.

- The $\top$ case is immediate because the checking and inference rules are the same; the identity substitution is returned from the inference algorithm.

- For the $m\ \oplus\ m'$ case, we are given two inductive hypotheses, $\forall\Gamma\varphi, \varphi\Gamma \Vdash m \to \varphi\Gamma \vdash m$ and $\forall\Gamma\varphi, \varphi\Gamma \Vdash m' \to \varphi\Gamma \vdash m'$. We must show that $\forall\Gamma\varphi,\ \varphi\Gamma \Vdash m\ \oplus\ m' \to \varphi\Gamma \vdash m\ \oplus\ m'$. Introduce Γ and φ and assume that $\varphi\Gamma \Vdash m\ \oplus\ m'$.

  By inversion, we know that there exists ϕ_1 and ϕ_2 such that $\varphi = \phi_2 \circ \phi_1$ and $\phi_1\Gamma \Vdash m$ and $\phi_2\phi_1\Gamma \Vdash m'$. Substitution yields a new goal of $\phi_2\phi_1\Gamma \vdash m\ \oplus\ m'$. We can apply the WF-ANDIMPL rule with $\phi_2\phi_1\Gamma$ as its Γ. This yields the two goals of $\phi_2\phi_1\Gamma \vdash m$ and $\phi_2\phi_1\Gamma \vdash m'$.

  To solve the first goal, by the fact that typechecking respects substitution (Lemma 3), we need to only solve $\phi_1\Gamma \vdash m$. We can then apply the inductive hypothesis using Γ and ϕ_1 to get a new goal of $\phi_1\Gamma \Vdash m$, we we assumed.

  To solve the second goal, we apply the inductive hypothesis using Γ and $\phi_2 \circ \phi_1$ to get a new goal of $\phi_2\phi_1\Gamma \Vdash m'$, which we assumed.

- For the $=$ case, we have no inductive hypotheses. Assume that $\varphi\Gamma \Vdash p_1 \ = \ p_2$; we must show that $\varphi\Gamma \vdash p_1 \ = \ p_2$. By inversion we know that there exists $\phi_4, \phi_3, \phi_2, \phi_1, t_1, t_2, \alpha$ such that $\varphi = \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$ and $\phi_1\Gamma \Vdash p_1 :: t_1$ and $t_1 \overset{\phi_2}{\approx} \mathtt{ATOMIC}\ \alpha$ and $\phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2$ and $\mathtt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\approx} t_2$. Substituting gives us a new goal of $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p_1 \ = \ p_2$.

  Using the soundness of inference for paths (Lemma 2) with $\phi_1\Gamma \Vdash p_1 :: t_1$ and $\phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2$ yields that $\phi_1\Gamma \vdash p_1 :: t_1$ and $\phi_3\phi_2\phi_1\Gamma \vdash p_2 :: t_2$. We can then repeatedly apply that path checking respects substitution (Lemma 1), to get (1) $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p_1 :: \phi_4\phi_3\phi_2 t_1$ and (2) $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p_2 :: \phi_4 t_2$.

The mgu property gives us that (a) $\phi_4 t_2 = \texttt{ATOMIC}\ \phi_4\phi_3\phi_2\alpha$ and (b) $\phi_2 t_1 = \phi_2\texttt{ATOMIC}\ \alpha$. Rewriting by (a) in (2) gives $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p_2 :: \texttt{ATOMIC}\ \phi_4\phi_3\phi_2\alpha$. Rewriting by (b) in (1) gives $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p_1 :: \texttt{ATOMIC}\ \phi_4\phi_3\phi_2\alpha$. Hence we can apply the WF-EQ rule.

- For the $\underline{\mathsf{in}}$ case, the inductive hypothesis is that $\forall\Gamma\varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$. We must show that $\forall\Gamma\varphi, \varphi\Gamma \Vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m \rightarrow \varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. To that end, introduce Γ and φ and assume that $\varphi\Gamma \Vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. By inversion, we know there exists $\phi_3, \phi_2, \phi_1, t, \rho$ with ρ fresh such that $\varphi = \phi_3 \circ \phi_2 \circ \phi_1$ and $\phi_1\Gamma \Vdash p :: t$ and $\texttt{SETRCD}\ \rho \overset{\phi_2}{\approx} t$ and $\phi_3\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma) \Vdash m$. We can then apply the SETRCD-ELIM rule, taking its Γ to be $\phi_3\phi_2\phi_1\Gamma$ and its r to be $\phi_3\phi_2\rho$. We are thus left with two new goals: $\phi_3\phi_2\phi_1\Gamma \vdash p :: \texttt{SETRCD}\ \phi_3\phi_2\rho$ and $(v, \texttt{RCD}\ \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \vdash m$.

  To solve the first, note that we have $\phi_1\Gamma \Vdash p :: t$ and thus by soundness of path inference we have that $\phi_1\Gamma \vdash p :: t$, and thus also that (1) $\phi_2\phi_1\Gamma \vdash p :: \phi_2 t$ because path checking respects substitution. By the mgu property, we have $\texttt{SETRCD}\ \phi_2\rho = \phi_2 t$, and so we can rewrite (1) to obtain $\phi_2\phi_1\Gamma \vdash p :: \texttt{SETRCD}\ \phi_2\rho$, and then apply that path checking respects substitution with ϕ_3 to get $\phi_3\phi_2\phi_1\Gamma \vdash p :: \texttt{SETRCD}\ \phi_3\phi_2\rho$, as required.

  To solve the second, apply the inductive hypothesis with its Γ as $(v, \texttt{SETRCD}\ \rho); \phi_1\Gamma$ and its φ as $\phi_3 \circ \phi_2$ to yield a new goal of $(v, \texttt{RCD}\ \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \Vdash m$, which we already assumed.

- For the $\underline{\mathsf{of}}$ case, the inductive hypothesis is that $\forall\Gamma\varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$. We must show that $\forall\Gamma\varphi, \varphi\Gamma \Vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m \rightarrow \varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. To that end, introduce Γ and φ and assume that $\varphi\Gamma \Vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. By inversion, we know there exists $\phi_3, \phi_2, \phi_1, t, \rho, \sigma$ with ρ, σ fresh such that $\varphi = \phi_3 \circ \phi_2 \circ \phi_1$ and $\phi_1\Gamma \Vdash p :: t$ and $\texttt{SETCHC}\ \llparenthesis l : \sigma, \rho \rrparenthesis \overset{\phi_2}{\approx} t$ and $\phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m$. We can then apply the SETCHC-ELIM rule, taking its Γ to be $\phi_3\phi_2\phi_1\Gamma$ and its r to be $\phi_3\phi_2\rho$ and its t to be $\phi_3\phi_2\sigma$. We are thus left with two new goals: $\phi_3\phi_2\phi_1\Gamma \vdash p :: \texttt{SETCHC}\ \llparenthesis l : \phi_3\phi_2\sigma, \phi_3\phi_2\rho \rrparenthesis$ and $(v, \texttt{RCD}\ \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \vdash m$.

  To solve the first, note that we have $\phi_1\Gamma \Vdash p :: t$ and thus by soundness of path inference we have that $\phi_1\Gamma \vdash p :: t$, and thus also that (1) $\phi_2\phi_1\Gamma \vdash p :: \phi_2 t$ because typing respects substitution. By the mgu property, we have $\texttt{SETCHC}\ \llparenthesis l : \phi_2\sigma, \phi_2\rho \rrparenthesis = \phi_2 t$, and so we can rewrite (1) to obtain $\phi_2\phi_1\Gamma \vdash p :: \texttt{SETCHC}\ \llparenthesis l : \phi_2\sigma, \phi_2\rho \rrparenthesis$, and then apply that path checking respects substitution with ϕ_3 to get $\phi_3\phi_2\phi_1\Gamma \vdash p : \texttt{SETCHC}\ \llparenthesis \phi_3\phi_2 l : \sigma, \phi_3\phi_2\rho \rrparenthesis$, as required.

  To solve the second, apply the inductive hypothesis with its Γ as $(v, \sigma); \phi_1\Gamma$ and its φ as $\phi_3 \circ \phi_2$ to yield a new goal of $(v, \phi_3\phi_2\sigma); \phi_3\phi_2\phi_1\Gamma \Vdash m$, which we already assumed.

□

Theorem 2 - Completeness

Intuitively, completeness (and soundness) holds because we are simply doing iterated unification in a way similar to ML's Hindley-Milner type inference or [3].

We start by establishing the completeness of path inference over path checking. Let $fv(\Gamma)$ indicate the union of the type, row, and atomic variables in the types in the range of context Γ. Write $s_1 =_\Gamma s_2$ to indicate that substitutions s_1 and s_2 are equal on variables appearing in Γ.

Lemma 5 (Completeness of Path Inference)**.**

$$\forall\ p\ \Gamma\ \tau\ \varphi,\ \varphi\Gamma \vdash p :: \tau \to \exists\ S\ T\ s,\ S\Gamma \Vdash p :: T \land \tau = sT \land \varphi =_\Gamma s \circ S$$

Proof. Proceed by induction on p.

- The base case is that p is a variable v, and we are to prove that (introducing Γ and τ and φ):

  $$\varphi\Gamma \vdash v :: \tau \to \exists\ S\ T\ s,\ S\Gamma \Vdash v :: T \land \tau =_\Gamma sT \land \varphi = s \circ S$$

  Assume $\varphi\Gamma \vdash v : \tau$. By inversion, there exists some T such that $(v, T) \in \Gamma$; witness S with the identity substitution, s with φ, and T with T. The goal is then

  $$\Gamma \Vdash v :: T \land \tau = \varphi T \land \varphi =_\Gamma \varphi$$

  The leftmost conjunct follows from the VAR-INF rule and that we have $(v, T) \in \Gamma$. The rightmost conjunct is trivial. The middle conjunct, $\tau = \varphi T$, follows because we have $(v, T) \in \Gamma$ and hence that $(v, \varphi T) \in \varphi\Gamma$. Because $(v, \tau) \in \varphi\Gamma$, it must be the case that $\tau = \varphi T$.

- The inductive case is that p is a projection $p.l$, and we are to prove that

  $$(\mathrm{G}OAL-1) \qquad \forall\ \Gamma\ \tau\ \varphi,\ \varphi\Gamma \vdash p.l : \tau \to \exists\ S\ T\ s,\ S\Gamma \Vdash p.l : T \land \tau = sT \land \varphi =_\Gamma s \circ S$$

  Introduce Γ and φ and τ and assume that

  $$(1) \qquad \varphi\Gamma \vdash p.l : \tau$$

  We can specialize the goal's existentials, taking $S = \psi \circ \phi$ and $T = \psi\sigma$ to obtain a goal of (where implicitly ρ, σ are fresh in the goal):

  $$(\mathrm{G}OAL-2) \qquad \exists\ \rho\ \sigma\ \psi\ \phi\ s', \psi\phi\Gamma \Vdash p.l : \psi\sigma \land \tau = s'\psi\sigma \land \varphi =_\Gamma s' \circ \psi \circ \phi$$

  By the RCD-ELIM-INF goal the goal is:

  $$(\mathrm{G}OAL-3) \qquad \exists\ \rho\ \sigma\ \psi\ \phi\ t\ s', \phi\Gamma \Vdash p : t \land \mathtt{RCD}\ (\!|l : \sigma, \rho|\!) \overset{\psi}{\sim} t \land \tau = s'\psi\sigma \land \varphi =_\Gamma s' \circ \psi \circ \phi$$

  The inductive hypothesis is that

  $$(\mathrm{I}H) \qquad \forall \Gamma\ \tau\ \varphi,\ \varphi\Gamma \vdash p : \tau \to \exists\ S\ T\ s,\ S\Gamma \Vdash p : T \land \tau = sT \land \varphi =_\Gamma s \circ S$$

  By inversion of (1), we know that there exists r such that

  $$(2) \qquad \varphi\Gamma \vdash p : \mathtt{RCD}\ (\!|l : \tau, r|\!)$$

Apply IH with $\Gamma = \Gamma$, $\tau = \texttt{RCD}\ \llparenthesis l : \tau, r \rrparenthesis$, and $\varphi = \varphi$ with (2) to obtain an S and a T and an s such that

$$(3) \qquad S\Gamma \Vdash p : T \wedge \texttt{RCD}\ \llparenthesis l : \tau, r \rrparenthesis = sT \wedge \varphi =_{\Gamma} s \circ S$$

We can then instantiate $\phi = S$ and $t = T$ to obtain a goal of

$$(\mathrm{G}OAL - 4) \qquad \exists\ \rho\ \sigma\ \psi\ s', S\Gamma \Vdash p : T \wedge \texttt{RCD}\ \llparenthesis l : \sigma, \rho \rrparenthesis \overset{\psi}{\sim} T \wedge \tau = s'\psi\sigma \wedge \varphi =_{\Gamma} s' \circ \psi \circ S$$

The leftmost conjunct we have in (3), to get a goal of

$$(\mathrm{G}OAL - 5) \qquad \exists\ \rho\sigma\ \psi\ s', \texttt{RCD}\ \llparenthesis l : \sigma, \rho \rrparenthesis \overset{\psi}{\sim} T \wedge \tau = s'\psi\sigma \wedge \varphi =_{\Gamma} s' \circ \psi \circ S$$

From (3), we know that $\texttt{RCD}\ \llparenthesis l : \tau, r \rrparenthesis = sT$, which means that T must have one of three forms:

- $T = u$, for some type variable u.
- $T = \texttt{RCD}\ \llparenthesis r', u \rrparenthesis$, for some row variable u and some row r' that does not contain label l.
- $T = \texttt{RCD}\ \llparenthesis l : \tau', r' \rrparenthesis$ for some row r' that does not contain label l.

In all cases there exists a ψ and $\sigma, \rho \notin dom(s)$ such that,

$$(4) \qquad T \overset{\psi}{\sim} \texttt{RCD}\ \llparenthesis l : \sigma, \rho \rrparenthesis$$

Witness our goal to get:

$$(\mathrm{G}OAL - 6) \qquad \exists\ s', \tau = s'\psi\sigma \wedge \varphi =_{\Gamma} s' \circ \psi \circ S$$

Rewrite $\varphi =_{\Gamma} s \circ S$ to get

$$(\mathrm{G}OAL - 7) \qquad \exists\ s', \tau = s'\psi\sigma \wedge s \circ S =_{\Gamma} s' \circ \psi \circ S$$

We know from (3) and because ρ, σ are fresh that $(\sigma \mapsto \tau, \rho \mapsto r) \circ s$ unifies $\texttt{RCD}\ \llparenthesis l : \sigma, \rho \rrparenthesis$ and T; hence there exists some z such that

$$(5) \qquad (\sigma \mapsto \tau, \rho \mapsto r) \circ s = z \circ \psi$$

Witnessing our goal with z gives a new goal of

$$(\mathrm{G}OAL - 8) \qquad \tau = z\psi\sigma \wedge s \circ S =_{\Gamma} z \circ \psi \circ S$$

Rewrite by (5) and the left goal disappears, because s cannot act on σ; get a goal of

$$(\mathrm{G}OAL - 9) \qquad s \circ S =_{\Gamma} (\sigma \mapsto \tau, \rho \mapsto r) \circ s \circ S$$

Which is true because $\sigma, \rho \notin fv(\Gamma)$.

□

The main theorem is

$$\forall\ m\ \Gamma\ \varphi,\ \ \varphi\Gamma \vdash m \rightarrow \exists\ S\ s,\ \ S\Gamma \Vdash m \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

Proof. By induction on m.

- For the case where $m = \top$, witness S as the identity substitution and s as φ. This yields a goal of $\Gamma \Vdash \top \wedge \varphi v = \varphi v$, which is trivially true.

- For the $m_1\ \oplus\ m_2$ case, we have two inductive hypotheses:

  $$\forall \Gamma\ \varphi, \varphi\Gamma \vdash m_1 \rightarrow \exists\ S\ s,\ \ S\Gamma \Vdash m_1 \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

  and

  $$\forall \Gamma\ \varphi, \varphi\Gamma \vdash m_2 \rightarrow \exists\ S\ s,\ \ S\Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

  Introduce Γ and φ and assume that $\varphi\Gamma \vdash m_1\ \oplus\ m_2$. By inversion, we know that $\varphi\Gamma \vdash m_1$ and that $\varphi\Gamma \vdash m_2$. We want to prove that

  $$\exists\ S\ s,\ \ S\Gamma \Vdash m_1\ \oplus\ m_2 \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

  We can tweak S to get:

  $$\exists\ \phi_1\ \phi_2\ s,\ \ \phi_2\phi_1\Gamma \Vdash m_1\ \oplus\ m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s\phi_2\phi_1 v$$

  And applying the AND-INF rule gives a new goal of

  $$\exists\ \phi_1\ \phi_2\ s,\ \ \phi_1\Gamma \Vdash m_1 \wedge \phi_2\phi_1\Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s\phi_2\phi_1 v$$

  Applying the first inductive hypothesis gives us a ϕ_1 and s_1 such that $\phi_1\Gamma \Vdash m_1 \wedge \forall v \in fv(\Gamma), \varphi v = s_1\phi_1 v$. We can then witness the goal and substitute for φ to get a new goal of

  $$\exists\ \phi_2\ s,\ \ \phi_2\phi_1\Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), s_1\phi_1 v = s\phi_2\phi_1 v$$

  Substituting for φ gives us $s_1\phi_1\Gamma \vdash m_2$. We can then apply the second inductive hypothesis (choosing its Γ to be $\phi_1\Gamma$ and its φ to be s_1, to obtain that there exists some ϕ_2 and s_2 such that $\phi_2\phi_1\Gamma \Vdash m_2 \wedge \forall v \in fv(\phi_1\Gamma), s_1 v = s_2\phi_2 v$. The goal's leftmost conjunct is then immediate, and the rightmost conjunct follows by inserting ϕ_1 on both sides of the substitution equality.

- For the $p_1\ =\ p_2$ case, we have no inductive hypothesis. Introduce Γ and φ and assume that $\varphi\Gamma \vdash p_1\ =\ p_2$. By inversion we know that there is some a such that $\varphi\Gamma \vdash p_1 :: \texttt{ATOMIC}\ a$ and $\varphi\Gamma \vdash p_2 :: \texttt{ATOMIC}\ a$. We want to prove that

  $$\exists\ S\ s,\ \ S\Gamma \Vdash p_1\ =\ p_2 \wedge \varphi =_\Gamma s \circ S$$

  Tweaking S yields a new goal of

  $$\exists\ \phi_1\ \phi_2\ \phi_3\ \phi_4\ s,\ \ \phi_4\phi_3\phi_2\phi_1\Gamma \Vdash p_1\ =\ p_2 \wedge \varphi =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  By the EQ-INF rule gives a new goal of

  $$\exists\ \alpha\ \phi_1\ \phi_2\ \phi_3\ \phi_4\ s\ t_1\ t_2,\ \ \phi_1\Gamma \Vdash p_1 :: t_1\ \wedge t_1 \overset{\phi_2}{\sim} \texttt{ATOMIC}\ \alpha\ \wedge\ \phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2 \wedge$$

  $$\texttt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\sim} t_2\ \wedge\ \varphi =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  (*) From completeness of path inference, we know that there exists some ϕ_1 and some t_1 and some s_1 such that $\phi_1\Gamma \Vdash p_1 :: t_1 \wedge \texttt{ATOMIC}\ a = s_1 t_1 \wedge \varphi =_\Gamma s_1 \circ \phi_1$. Witnessing the goal, substituting for φ, and eliminating the $\phi_1\Gamma \Vdash p_1 :: t_1$ conjunct (since we now have it), gives the new goal

$$\exists\ \alpha\ \phi_2\ \phi_3\ \phi_4\ s\ t_2,\ t_1 \overset{\phi_2}{\sim} \texttt{ATOMIC}\ \alpha\ \wedge\ \phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2\ \wedge$$

$$\texttt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\sim} t_2\ \wedge\ s_1 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

Choose an α fresh for s_1 and t_1 and consider that $(\alpha \mapsto a) \circ s_1$ unifies $\texttt{ATOMIC}\ \alpha$ and t_1; that is, consider that

$$(\alpha \mapsto a)s_1(\texttt{ATOMIC}\ \alpha) = (\alpha \mapsto a)s_1t_1$$

is the same as

$$\texttt{ATOMIC}\ a = s_1t_1$$

Which we already know. This is because α is fresh for s_1; hence s_1 cannot act on it; neither can α appear in s_1t_1. Since there is a unifier, we know there must be a mgu ϕ_2, and so we witness the goal and remove a conjunct to get a goal of:

$$\exists\ \phi_3\ \phi_4\ s\ t_2,\ \phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2\ \wedge\ \texttt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\sim} t_2\ \wedge\ s_1 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

By the mgu property we know that there exists an s_2 such that

$$(\alpha \mapsto a) \circ s_1 = s_2 \circ \phi_2$$

Substituting into the goal (which is ok because α is not free in Γ) gives a goal of

$$\exists\ \phi_3\ \phi_4\ s\ t_2,\ \phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2\ \wedge\ \texttt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\sim} t_2\ \wedge\ s_2 \circ \phi_2 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

Substituting for φ into (*) gives that $s_1\phi_1\Gamma \vdash p_2 :: \texttt{ATOMIC}\ a$ Substituting for s_1 gives $s_2\phi_2\phi_1\Gamma \vdash p_2 :: \texttt{ATOMIC}\ a$. Hence by completeness of path inference (taking its Γ to be $\phi_2\phi_1\Gamma$ and its φ to be s_2) there is some ϕ_3 and t_2 and s_3 such that $\phi_3\phi_2\phi_1\Gamma \Vdash p_2 :: t_2 \wedge \texttt{ATOMIC}\ a = s_3t_2 \wedge s_2 =_{\phi_2\phi_1\Gamma} s_3 \circ \phi_3$. The goal now becomes (again eliminating a proved conjunct and substituting for s_2 and witnessing the existentials):

$$\exists\ \phi_4\ s, \texttt{ATOMIC}\ \phi_3\phi_2\alpha \overset{\phi_4}{\sim} t_2\ \wedge\ s_3 \circ \phi_3 \circ \phi_2 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

Note that s_3 unfies $\texttt{ATOMIC}\ \phi_3\phi_2\alpha$ and t_2. That is, consider

$$s_3(\texttt{ATOMIC}\ \phi_3\phi_2\alpha) = s_3t_2$$

The RHS is equal to $\texttt{ATOMIC}\ a$, and recall that $s_3 \circ \phi_3 = s_2$, and so the above is equivalent to

$$s_2(\texttt{ATOMIC}\ \phi_2\alpha) = \texttt{ATOMIC}\ a$$

But we also know that $s_2 \circ \phi_2 = (\alpha \mapsto a) \circ s_1$, and so the above is equivalent to

$$\texttt{ATOMIC}\ a = \texttt{ATOMIC}\ a$$

So, since there is a unifier there must be an mgu ϕ_4, and so the goal becomes

$$\exists\ s, s_3 \circ \phi_3 \circ \phi_2 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

Since ϕ_4 is the mgu, there must exist an s such that $s_3 = s \circ \phi_4$, and we can substitute and witness into the goal to get

$$s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1 =_\Gamma s \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

And we're done.

- When $m = \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$, we want to prove that

  $$\forall\ \Gamma\ \varphi,\ \ \varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m \rightarrow \exists\ S\ s,\ \ S\Gamma \Vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m\ \ \wedge \varphi =_{\Gamma} s \circ S$$

  Introduce Γ and φ and assume that $\varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. (*) By inversion there is some r such that $\varphi\Gamma \vdash p :: \texttt{SETRCD}\ r$, and that $(v, \texttt{RCD}\ r); \varphi\Gamma \vdash m$.

  Tweak the goal $S = \phi_3 \circ \phi_2 \circ \phi_1$ to get a goal of

  $$\exists\ \phi_3\ \phi_2\ \phi_1\ s,\ \ \phi_3\phi_2\phi_1\Gamma \Vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m\ \ \wedge \varphi =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Apply SETRCD-ELIM-INF to get a goal of

  $$\exists\ \rho\ t\ \phi_3\ \phi_2\ \phi_1\ s,\ \ \phi_1\Gamma \Vdash p :: t \wedge \texttt{SETRCD}\ \rho \overset{\phi_2}{\sim} t \wedge \phi_3\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma) \Vdash m\ \ \wedge \varphi =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  By completeness of path inference we have that there exists ϕ_1, T, s_1 such that

  $$\phi_1\Gamma \Vdash p :: T \wedge \texttt{SETRCD}\ r = s_1 T \wedge \varphi =_{\Gamma} s_1 \circ \phi_1$$

  Witness the goal and substitute for φ and removing the proven leftmost conjunct gives the goal:

  $$\exists\ \rho\ \phi_3\ \phi_2\ s,\ \ \texttt{SETRCD}\ \rho \overset{\phi_2}{\sim} T \wedge \phi_3\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma) \Vdash m\ \ \wedge s_1 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Choose a ρ fresh for s_1 and T, and consider that $(\rho \mapsto r) \circ s_1$ unifies $\texttt{SETRCD}\ \rho$ and T; that is,

  $$(\rho \mapsto r)s_1(\texttt{SETRCD}\ \rho) = (\rho \mapsto r)s_1 T$$

  because s_1 cannot act on ρ, since ρ is fresh for s_1; neither can ρ appear in $s_1 T$; we're left with $\texttt{SETRCD}\ r = s_1 T$, which we already know. Because there is a unifier, there must be an mgu ϕ_2, and the goal is now:

  $$\exists\ \phi_3\ s,\ \ \phi_3\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma) \Vdash m\ \ \wedge s_1 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Moreover, by the mgu property there is an s_2 such that

  $$(\rho \mapsto r) \circ s_1 = s_2 \circ \phi_2$$

  Substituting in for s_1 in the goal (which is possible because $\rho \notin fv(\Gamma)$), gives

  $$\exists\ \phi_3\ s,\ \ \phi_3\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma) \Vdash m\ \ \wedge s_2 \circ \phi_2 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  The inductive hypothesis is

  $$\forall\ \Gamma\ \varphi,\ \ \varphi\Gamma \vdash m \rightarrow \exists\ S\ s,\ \ S\Gamma \Vdash m\ \ \wedge \varphi =_{\Gamma} s \circ S$$

  Apply the IH with $\Gamma = \phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma)$, $\varphi = s_2$,

  $$s_2(\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma)) \vdash m \rightarrow \exists\ \phi_3\ s,\ \ \phi_3(\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma)) \Vdash m\ \wedge s_2 =_{\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma)} s \circ \phi_3$$

  Recall from the mgu discussion that $s_2 \circ \phi_2$ will send $\texttt{RCD}\ \rho$ to $\texttt{RCD}\ r$, so the antecedent above is $(v, \texttt{RCD}\ r); s_2\phi_2\phi_1\Gamma \vdash m$.

Recall from (*) that $(v, \texttt{RCD}\ r); \varphi\Gamma \vdash m$; we also know that $\varphi =_{\Gamma} s_2 \circ \phi_2 \circ \phi_1$. Hence we have the antecedent and so there must exist ϕ_3 and s such that

$$\phi_3(\phi_2((v, \texttt{RCD}\ \rho); \phi_1\Gamma)) \Vdash m \ \wedge s_2 =_{\phi_2((v,\texttt{RCD}\ \rho);\phi_1\Gamma)} s \circ \phi_3$$

Witness these into our goal; the leftmost of these applies directly, leaving a goal of

$$s_2 \circ \phi_2 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

This follows from the rightmost conjunct above.

- When $m = \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$, we want to prove that

  $$\forall\ \Gamma\ \varphi,\ \varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m \rightarrow \exists\ S\ s,\ S\Gamma \Vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ \ p.\ m\ \wedge \varphi =_{\Gamma} s \circ S$$

  Introduce Γ and φ and assume that $\varphi\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. (*) By inversion there is some t and r such that $\varphi\Gamma \vdash p :: \texttt{SETCHC}\ \llparenthesis l : t, r \rrparenthesis$, and that $(v, t); \varphi\Gamma \vdash m$.

  Tweak the goal $S = \phi_3 \circ \phi_2 \circ \phi_1$ to get a goal of

  $$\exists\ \phi_3\ \phi_2\ \phi_1\ s,\ \phi_3\phi_2\phi_1\Gamma \Vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m\ \wedge \varphi =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Apply SETRCD-ELIM-INF to get a goal of

  $$\exists\, \rho\, \sigma\, t\, \phi_3\ \phi_2\ \phi_1\ s,\ \phi_1\Gamma \Vdash p :: t \wedge \texttt{SETCHC}\ \llparenthesis l : \sigma, \rho \rrparenthesis \overset{\phi_2}{\approx} t \wedge \phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m \wedge \varphi =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  By completeness of path inference we have that there exists ϕ_1, T, s_1 such that

  $$\phi_1\Gamma \Vdash p :: T \wedge \texttt{SETCHC}\ \llparenthesis l : t, r \rrparenthesis = s_1 T \wedge \varphi =_{\Gamma} s_1 \circ \phi_1$$

  Witness the goal and substitute for φ and removing the proven leftmost conjunct gives the goal:

  $$\exists\ \rho\ \sigma\ \phi_3\ \phi_2\ s,\ \texttt{SETCHC}\ \llparenthesis l : \sigma, \rho \rrparenthesis \overset{\phi_2}{\approx} T \wedge \phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m\ \wedge s_1 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Choose a ρ and σ fresh for s_1 and T, and consider that $(\rho \mapsto r, \sigma \mapsto t) \circ s_1$ unifies $\texttt{SETCHC}\ \llparenthesis l : \sigma, \rho \rrparenthesis$ and T; that is,

  $$(\rho \mapsto r, \sigma \mapsto t)s_1(\texttt{SETCHC}\ \llparenthesis l : \sigma, \rho \rrparenthesis) = (\rho \mapsto r, \sigma \mapsto t)s_1 T$$

  because s_1 cannot act on ρ and σ, since they are fresh for s_1; neither can ρ or σ appear in $s_1 T$; we're left with $\texttt{SETCHC}\ \llparenthesis l : t, r \rrparenthesis = s_1 T$, which we already know. Because there is a unifier, there must be an mgu ϕ_2, and the goal is now:

  $$\exists\ \phi_3\ s,\ \phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m\ \wedge s_1 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

  Moreover, by the mgu property there is an s_2 such that

  $$(\rho \mapsto r, \sigma \mapsto t) \circ s_1 = s_2 \circ \phi_2$$

  Substituting in for s_1 in the goal (which is possible because $\sigma, \rho \notin fv(\Gamma)$), gives

  $$\exists\ \phi_3\ s,\ \phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m\ \wedge s_2 \circ \phi_2 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

The inductive hypothesis is

$$\forall\ \Gamma\ \varphi,\ \varphi\Gamma \vdash m \rightarrow \exists\ S\ s,\ S\Gamma \Vdash m\ \ \wedge \varphi =_{\Gamma} s \circ S$$

Apply the IH with $\Gamma = \phi_2((v,\sigma);\phi_1\Gamma)$, $\varphi = s_2$,

$$s_2(\phi_2((v,\sigma);\phi_1\Gamma)) \vdash m \rightarrow \exists\ \phi_3\ s,\ \phi_3(\phi_2((v,\sigma);\phi_1\Gamma)) \Vdash m\ \ \wedge s_2 =_{\phi_2((v,\sigma);\phi_1\Gamma)} s \circ \phi_3$$

Recall from the mgu discussion that $s_2 \circ \phi_2$ will send σ to t, so the antecedent above is $(v,t); s_2\phi_2\phi_1\Gamma \vdash m$.

Recall from (*) that $(v,t);\varphi\Gamma \vdash m$; we also know that $\varphi =_{\Gamma} s_2 \circ \phi_2 \circ \phi_1$. Hence we have the antecedent and so there must exist ϕ_3 and s such that

$$\phi_3(\phi_2((v,t);\phi_1\Gamma)) \Vdash m\ \ \wedge s_2 =_{\phi_2((v,t);\phi_1\Gamma)} s \circ \phi_3$$

Witness these into our goal; the leftmost of these applies directly, leaving a goal of

$$s_2 \circ \phi_2 \circ \phi_1 =_{\Gamma} s \circ \phi_3 \circ \phi_2 \circ \phi_1$$

This follows from the rightmost conjunct above.

□

Theorem 4 - Mapping Subtyping

We want to prove that

$$\forall\ m\ \Gamma'\ \Gamma,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

First, an auxillary lemma about path checking.

Lemma 6. $\forall\ p\ \Gamma'\ \Gamma\ t,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash p :: t \rightarrow \exists t', \Gamma' \vdash p :: t' \wedge t' \leq t$.

Proof. Proceed by induction on p.

- The base case is that p is a variable v. Introduce Γ' and Γ and t and assume that $\Gamma' \leq \Gamma$ and $\Gamma \vdash v :: t$. We want to show that $\exists t', \Gamma' \vdash v :: t' \wedge t' \leq t$. By inversion we know that $(v, t) \in \Gamma$ and because $\Gamma' \leq \Gamma$ that there exists some t' such that (A) $t' \leq t$ and (B) $(v, t') \in \Gamma'$. Witness the goal with t', so we must show that $t' \leq t$ and $\Gamma' \vdash v :: t'$. The first of these we already have in (A), and the second follows by applying the VAR rule with (B).
- The inductive case is that p is a projection $p.l$. The inductive hypothesis is that

  $$\forall \Gamma'\ \Gamma\ t,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash p :: t \rightarrow \exists t', \Gamma' \vdash p :: t' \wedge t' \leq t$$

  Introduce Γ' and Γ and t, and assume that $\Gamma' \leq \Gamma$ and $\Gamma \vdash p.l :: t$. We are to prove that

  $$\exists t', \Gamma' \vdash p.l :: t' \wedge t' \leq t$$

  By inversion, we know that there exists r such that $\Gamma \vdash p :: \texttt{RCD}\ (\!|l : t, r|\!)$. Apply the inductive hypothesis to get that there exists t' such that

  $$(C) \quad \Gamma' \vdash p :: t' \wedge t' \leq \texttt{RCD}\ (\!|l : t, r|\!)$$

  From the right conjunct we know that there exists T and R such that

  $$(D) \quad t' = \texttt{RCD}\ (\!|l : T, R|\!) \wedge T \leq t \wedge R \leq r$$

  Witness our goal with T so that we need to prove

  $$\Gamma' \vdash p.l :: T \wedge T \leq t$$

  The rightmost conjunct follows from (D). We can rewrite (C) by (D) to obtain

  $$\Gamma' \vdash p :: \texttt{RCD}\ (\!|l : T, R|\!)$$

  from which our goal follows by the RCD-ELIM rule.

□

Now the main result.

Proof. Proceed by induction on m.

- For the case where $m = \top$, note that $\top$ typechecks in any context, and we're done.
- For the case where $m = m_1\ \oplus\ m_2$, we have two inductive hypothesis, that

  $$\forall \Gamma'\ \Gamma,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash m_1 \rightarrow \Gamma' \vdash m_1$$

  and

  $$\forall \Gamma'\ \Gamma,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash m_2 \rightarrow \Gamma' \vdash m_2$$

  Introduce Γ' and Γ and assume that $\Gamma' \leq \Gamma$ and $\Gamma \vdash m_1\ \oplus\ m_2$. We want to prove that

  $$\Gamma' \vdash m_1\ \oplus\ m_2$$

  By inversion we know that $\Gamma \vdash m_1$ and $\Gamma \vdash m_2$, which we can use with the inductive hypothesis to obtain that $\Gamma' \vdash m_1$ and $\Gamma' \vdash m_2$. Our goal follows by the WF-ANDIMPL rule.

- For the case where $m = p_1 \;=\; p_2$, we have no inductive hypothesis. Introduce Γ' and Γ and assume that $\Gamma' \leq \Gamma$ and $\Gamma \vdash p_1 \;=\; p_2$. We want to prove that

  $$\Gamma' \vdash p_1 \;=\; p_2$$

  By inversion we know that there is some a such that $\Gamma \vdash p_1 :: \mathtt{ATOMIC}\ a$ and $\Gamma \vdash p_2 :: \mathtt{ATOMIC}\ a$. Using these with the lemma we just proved gives us that there exists $T_1 \leq \mathtt{ATOMIC}\ a$ and $T_2 \leq \mathtt{ATOMIC}\ a$ such that $\Gamma' \vdash p_1 :: T_1$ and $\Gamma' \vdash p_2 :: T_2$. However, T_1 and T_2 must be equal to $\mathtt{ATOMIC}\ a$, because reflexivity is the only subtyping rule that applies to atomic types. The goal then follows from the WF-EQ rule.

- For the $\diamond\ v\ \underline{\mathsf{in}}\ p.\ m$ case, the inductive hypothesis is

  $$\forall \Gamma'\ \Gamma,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

  Introduce Γ' and Γ and assume that $\Gamma' \leq \Gamma$ and that $\Gamma \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$. We want to show that

  $$\Gamma' \vdash \diamond\ v\ \underline{\mathsf{in}}\ p.\ m$$

  By inversion, we know that there is some r such that $(v, \mathtt{RCD}\ r); \Gamma \vdash m$ and $\Gamma \vdash p :: \mathtt{SETRCD}\ r$. From the lemma we just proved we have that there is some $R \preceq r$ such that (A) $\Gamma' \vdash p :: \mathtt{SETRCD}\ R$, and so we can apply the inductive hypothesis taking its Γ to be $(v, \mathtt{RCD}\ r); \Gamma$ and its Γ' to be $(v, \mathtt{RCD}\ R); \Gamma'$ to obtain

  $$(B) \quad (v, \mathtt{RCD}\ R); \Gamma' \vdash m$$

  From which the result follows by the SETRCD-ELIM rule applied with (A) and (B).

- For the $\diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$ case, the inductive hypothesis is

  $$\forall \Gamma'\ \Gamma,\ \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

  Introduce Γ' and Γ and assume that $\Gamma' \leq \Gamma$ and that $\Gamma \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$. The goal is

  $$\Gamma' \vdash \diamond\ v\ \underline{\mathsf{of}}\ l\ \underline{\mathsf{from}}\ p.\ m$$

  By inversion, there is some t and r such that $(v, t); \Gamma \vdash m$ and $\Gamma \vdash p :: \mathtt{SETCHC}\ \llparenthesis l : t, r \rrparenthesis$. From the lemma we just proved we have that there is some $R \preceq r$ and $T \leq t$ such that (A) $\Gamma' \vdash p :: \mathtt{SETCHC}\ \llparenthesis l : T, R \rrparenthesis$, and so we can apply the inductive hypothesis taking its Γ to be $(v, t); \Gamma$ and its Γ' to be $(v, T); \Gamma'$ to obtain

  $$(B) \quad (v, T); \Gamma' \vdash m$$

  From which the result follows by the SETCHC-ELIM rule applied with (A) and (B).

□

Proposition 3

Suppose we have that $X' < X$ and $I \in [\![X']\!]$. Intuitively, the derivation used for *erase* does not matter because the fields removed from the data instance are exactly those found in X' but not in X – which does not depend on the order in which we determine these fields.

Proof. More formally, let the rank of a schema be the maximum number of *Row* nestings, so that NR schema are thought of as trees. The proof is by induction on the rank of X'. Rank induction says that for a predicate $P(X')$ of a schema X', if we can establish $P(X')$ for all schema of rank 0, and assuming that $P(X')$ holds for all schema of rank less than n, that $P(X')$ holds of all schema of rank n, then we may conclude that $P(X')$ holds for all schema. The predicate we want to use is

$$P(X') = \forall\ X\ (I \in [\![X']\!])\ (pf\ pf' : X' < X),\ erase(pf)(I) = erase(pf')(I)$$

For the base case, when the rank is zero, X' is an `ATOMIC`, and the subtyping derivations are both necessarily uses of reflexivity, which removes no data from I.

For the inductive step, suppose the rank of X' is n and the inductive hypothesis is that we've established the irrelevence of the subtyping derivation for all ranks less than n. X' is of the form Cr' for some row r', and X has the form Cr for some r such that $r' \prec r$.

Note that for each $(l : t) \in r'$, that the rank of t is less than n. The effect of $erase(pf)$ and $erase(pf')$ must necessarily consist of a sequence of width-removals and depth-removals to the labels in I. For each l that occurs in X' but not in X, the final erasure to apply to that label must be width; hence, we just need to guarantee that the order in which the depth erasures were applied to labels that do appear in the output does not matter. In this case, it must be the case that the final erasure to apply to a label erases into the label's type in X; as such, the inductive hypothesis guarantees that the particular derivation used to do this erasure does not matter.

□

Theorem 5

First, a lemma.

Lemma 7.

$$\forall\, p\, \Gamma'\, \Gamma\, t'\, t\, I'\, \Delta',\ \Gamma' \leq \Gamma \ \wedge\ t' \leq t \ \wedge\ I' \in [\![t']\!] \ \wedge\ \Delta' \in [\![\Gamma']\!] \ \wedge \Gamma \vdash p :: t \wedge \Gamma' \vdash p :: t' \ \wedge\ \Delta' \models p \rightsquigarrow I' \rightarrow$$

$$erase(\Gamma' \leq \Gamma)(\Delta') \models p \rightsquigarrow erase(t' \leq t)(I')$$

Proof. By induction on p

- When p is a variable v, we assume that

  $$\Gamma' \leq \Gamma \ \wedge\ t' \leq t \ \wedge\ I' \in [\![t']\!] \ \wedge\ \Delta' \in [\![\Gamma']\!] \ \wedge \Gamma \vdash v :: t \wedge \Gamma' \vdash v :: t' \ \wedge\ \Delta' \models v \rightsquigarrow I'$$

  and we must show

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models v \rightsquigarrow erase(t' \leq t)(I')$$

  This follows because $(v, t') \in \Gamma'$ and $(v, t) \in \Gamma$ and $(v, I') \in \Delta'$.

- For the inductive step where $p = p.l$, we assume that

  $$\Gamma' \leq \Gamma \ \wedge\ t' \leq t \ \wedge\ I' \in [\![t']\!] \ \wedge\ \Delta' \in [\![\Gamma']\!] \ \wedge \Gamma \vdash p.l :: t \wedge \Gamma' \vdash p.l :: t' \ \wedge\ \Delta' \models p.l \rightsquigarrow I'$$

  and want to show that

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models p.l \rightsquigarrow erase(t' \leq t)(I')$$

  By inversion we know that

  $$\Gamma \vdash p :: \texttt{RCD}\ (\!|l : t, r|\!) \wedge \Gamma' \vdash p :: \texttt{RCD}\ (\!|l : t', r'|\!) \ \wedge$$

  $$\Delta' \models p \rightsquigarrow \{(l : I')\} \cup X$$

  where $X \in [\![\texttt{RCD}\ r']\!]$. The inductive hypothesis is

  $$\forall\, \Gamma'\, \Gamma\, t'\, t\, I'\, \Delta',\ \Gamma' \leq \Gamma \wedge t' \leq t \wedge I' \in [\![t']\!] \wedge \Delta' \in [\![\Gamma']\!] \wedge \Gamma \vdash p :: t \wedge \Gamma' \vdash p :: t' \wedge \Delta' \models p \rightsquigarrow I' \rightarrow$$

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models p \rightsquigarrow erase(t' \leq t)(I')$$

  Which we apply with $t' = \texttt{RCD}\ (\!|l : t', r'|\!)$ and $t = \texttt{RCD}\ (\!|l : t, r|\!)$ and $I' = \{(l : I')\} \cup X$ to obtain

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models p \rightsquigarrow erase(\texttt{RCD}\ (\!|l : t', r'|\!) \leq \texttt{RCD}\ (\!|l : t, r|\!))(\{(l : I')\} \cup X)$$

  which simplifies to

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models p \rightsquigarrow \{(l : erase(t' \leq t)(I'))\} \cup erase(\texttt{RCD}\ r' \leq \texttt{RCD}\ r)(X)$$

  And hence

  $$erase(\Gamma' \leq \Gamma)(\Delta') \models p.l \rightsquigarrow erase(t' \leq t)(I')$$

  As required.

□

Now for the theorem itself.

Suppose $\Gamma \vdash M$ and $\Gamma' \leq \Gamma$ and $\Delta' \in [\![\Gamma']\!]$. Then $\Delta' \models M$ if and only if $\mathit{erase}(\Gamma' \leq \Gamma)\ (\Delta') \models M$.

Proof. We prove by induction on $\Gamma \vdash M$ that whenever $\Gamma \vdash M$, $\Gamma' \leq \Gamma$ and $\Delta' \in [\![\Gamma']\!]$ then $\Delta' \models M$ if and only if $\mathit{erase}(\Gamma' \leq \Gamma)\ (\Delta') \models M$. We show here the case for $\underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$. The other cases are similar or simpler.

Let $\Gamma \vdash \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$. Then it must be the case, by the SETRCD-ELIM rule, that $\Gamma \vdash p :: \mathtt{SETRCD}\ r$ and $(v, \mathtt{RCD}\ r); \Gamma \vdash M$. Let $\Gamma' \leq \Gamma$, $\Delta' \in [\![\Gamma']\!]$ and $\Delta = \mathit{erase}(\Gamma' \leq \Gamma)\ (\Delta')$.

It is easy to see that $\Gamma' \leq \Gamma$ and $\Gamma \vdash p :: \mathtt{SETRCD}\ r$ implies that $\Gamma' \vdash p :: \mathtt{SETRCD}\ r'$ where $r' \prec r$ (and hence $\mathtt{RCD}\ r' \leq \mathtt{RCD}\ r$). From $\mathtt{RCD}\ r' \leq \mathtt{RCD}\ r$ and $\Gamma' \leq \Gamma$, it follows that $(v, \mathtt{RCD}\ r'); \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma$. By Theorem 4, since $(v, \mathtt{RCD}\ r); \Gamma \vdash M$, we obtain that $(v, \mathtt{RCD}\ r'); \Gamma' \vdash M$.

We now show that $\Delta' \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$ if and only if $\Delta \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$.

Let us assume now that $\Delta' \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$. We must show that $\Delta \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$. Since $\Delta' \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$ we must have that $\Delta' \models p \rightsquigarrow I'$, where $I' \in [\![\mathtt{SETRCD}\ r']\!]$, and $\forall i' \in I'$, $(v, i'); \Delta' \models M$.

Applying the lemma, we obtain that $\Delta \models p \rightsquigarrow I$, where $I = \mathit{erase}(\mathtt{SETRCD}\ r' \leq \mathtt{SETRCD}\ r)\ (I')$. We also have (by the definition of *erase* for SETRCD) that for every $i \in I$ there is some $i' \in I'$ such that $i = \mathit{erase}(\mathtt{RCD}\ r' \leq \mathtt{RCD}\ r)\ (i')$. We also know that i' satisfies $(v, i'); \Delta' \models M$ (this is true for every $i' \in I'$, by the earlier fact). We are now ready to apply the inductive hypothesis, which is:

$$\forall\ \Gamma'\ \Delta' \in [\![\Gamma']\!], \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma \rightarrow (\Delta' \models M \iff \mathit{erase}(\Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma)(\Delta') \models M)$$

We have the following facts already established: $(v, \mathtt{RCD}\ r); \Gamma \vdash M$, $(v, \mathtt{RCD}\ r'); \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma$, $(v, i'); \Delta' \models M$. Applying the inductive hypothesis (in one direction) taking its Γ' to be $(v, \mathtt{RCD}\ r'); \Gamma'$ and its Δ' to be $(v, i'); \Delta'$ yields

$$\mathit{erase}((v, \mathtt{RCD}\ r'); \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma)((v, i'); \Delta') \models M.$$

But the above environment (on the left-hand side of $\models$) equals

$$\mathit{erase}((v, \mathtt{RCD}\ r'); \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma)((v, i')); \mathit{erase}((v, \mathtt{RCD}\ r'); \Gamma' \leq (v, \mathtt{RCD}\ r); \Gamma)(\Delta'),$$

which equals

$$(v, \mathit{erase}(\mathtt{RCD}\ r' \leq \mathtt{RCD}\ r)(i')); \mathit{erase}(\Gamma' \leq \Gamma)(\Delta'),$$

which, by earlier facts, equals $(v, i); \Delta$. Thus, we proved that $(v, i); \Delta \models M$, for every i in I. Together with the fact that $\Delta \models p \rightsquigarrow I$, we obtain that $\Delta \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$, which was to be proven.

The proof of the converse direction, that is, $\Delta \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$ implies $\Delta' \models \underline{\mathsf{for}}\ v\ \underline{\mathsf{in}}\ p.\ M$, is similar and makes use of the other direction of the inductive hypothesis.

□

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison Wesley Publishing Co, 1995.

[2] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[3] B. R. Gaster and M. P. Jones. A polymorphic type system for extensible records and variants. Technical Report NOTTCS-TR-96-3, Department of Computer Science, University of Nottingham, November 1996.