Completeness of E-Unification with Eager Variable Elimination

Barbara Morawska

Theoretical Computer Science, TU Dresden, D-01062 Dresden, Germany,
morawska@tcs.inf.tu-dresden.de

Abstract. The paper contains a proof of completeness of a goal-directed inference system for general E-unification with eager Variable Elimination. The proof is based on a careful analysis of a concept of ground, equational proof. The theory of equational proofs is developed in the first part. Solving variables in a goal is then shown to be reflected in defined transformations of an equational proof. The termination of these transformations proves termination of inferences with eager Variable Elimination.

1 Introduction

E-unification is concerned with finding a set of solutions for a given equation in a given equational theory E. The problem of E-unification arises in many areas of computer science like formal verification, theorem proving and logic programming. In general the E-unification problem, i.e. the problem of finding a set of solutions for a given equation in a non-empty equational theory E is undecidable, unlike in the case of the syntactic unification problem, i.e. in the case of searching for a solution for an equation in the context of the empty equational theory. Nevertheless, the E-unification problem is semi-decidable and there are complete algorithms for solving it.

Goal-directed algorithms for E-unification are based on the idea of transforming goal equations into a solved form which will allow easily to define a solution. Such an inference system was presented first in [2], and is displayed here in a different notation in Figure 1. Consider the rule Variable Elimination in this set of inference rules. If applied to an equation of the form $x \approx v$ in the goal, it will eliminate x from all other equations in the goal and thus solve the equation $x \approx v$.[1] The Variable Elimination is forced to be applied eagerly here, because there is no other rule to deal with equations of the form $x \approx v$, where x is not a variable in v.

There was no proof up to now that this system of inferences (Figure 1, page 199) is complete for E-unification. It is complete, when we allow other rules to apply to an equation $x \approx v$, but then Variable Elimination cannot be applied eagerly. The problem was first discovered and called *the Eager Variable Elimination Problem* by Gallier and Snyder in [2].

[1] Formal definition of a solved equation is in the section 6.

M.Y. Vardi and A. Voronkov (Eds.): LPAR 2003, LNAI 2850, pp. 198–212, 2003.

Decomposition

$$\frac{\{f(s_1,\cdots,s_n) \approx f(t_1,\cdots,t_n)\} \cup G}{\{s_1 \approx t_1,\cdots,s_n \approx t_n\} \cup G}$$

where $f(s_1,\cdots,s_n) \approx f(t_1,\cdots,t_n)$ is selected in the goal.

Mutate

$$\frac{\{u \approx f(v_1,\cdots,v_n)\} \cup G}{\{u \approx s, t_1 \approx v_1,\cdots,t_n \approx v_n\} \cup G}$$

where $u \approx f(v_1,\cdots,v_n)$ is selected in the goal, and $s \approx f(t_1,\cdots,t_n) \in E$.[a]

Variable Mutate

$$\frac{\{u \approx f(v_1,\cdots,v_n)\} \cup G}{\{u \approx s, x \approx f(v_1,\cdots,v_n)\} \cup G}$$

where $s \approx x \in E$, x is a variable, and $u \approx f(v_1,\cdots,v_n)$ is selected in the goal.

Variable Decomposition (for cycle)

$$\frac{\{x \approx f(t_1,\cdots,t_n)\} \cup G}{\{x \approx f(x_1,\cdots,x_n)\} \cup (\{x_1 \approx t_1,\cdots,x_n \approx t_n\} \cup G)[x \mapsto f(x_1,\cdots,x_n)]}$$

where x is a variable, $x \approx f(t_1,\cdots,t_n)$ is selected in the goal,
$x \in Var(f(t_1,\cdots,t_n))$.

Variable Elimination

$$\frac{\{x \approx v\} \cup G}{\{x \approx v\} \cup G[x \mapsto v]}$$

where $x \notin Var(v)$

Orient

$$\frac{\{t \approx x\} \cup G}{\{x \approx t\} \cup G}$$

where x is a variable.
and t is not a variable.

Trivial

$$\frac{\{x \approx x\} \cup G}{G}$$

where $x \approx x$ is selected in the goal.

[a] We assume that E is closed under symmetry.

Fig. 1. E-Unification with eager Variable Elimination

Eager Variable Elimination is justified in the context of syntactic unification (the empty equational theory) because it decreases the number of unsolved variables in the goal, while preserving a set of solutions. The number of unsolved variables is not increased by any other rule and hence we may be sure that the inferences will terminate.

In the context of E-unification we must use another rule called here Mutate.[2] Notice that we have here conflicting results of applications of Mutate and Variable Elimination to the goal: Variable Elimination decreases the number of unsolved variables in the goal, but Mutate increases this number, and while Mutate decreases the length of a ground proof of an instance of a goal, Variable Elimination may increase this length.

In [3] Gallier and Snyder proved completeness of their system without eager Variable Elimination. In [5] the authors stated that Mutate (*replacement*) and eager Variable Elimination (*merging*) do not preserve the form of the proof.

In this paper we prove that Variable Elimination may be applied eagerly without destroying the completeness of the E-unification procedure. The fact that Variable Elimination can be applied eagerly decreases non-determinism in the inherently non-deterministic general E-unification algorithms. It may reduce redundancy of inferences and limit the search space for a solution to a given equation. This was pointed out e.g. in [7], [6], [4].

The main idea in the proof of completeness of our inference rules (Figure 1), is to consider a ground equational proof for a goal. Most of this paper is concerned with a description of a theory of equational ground proofs (definitions in Section 3) and a construction of new equational proof which reflects effects of eager Variable Elimination (Section 4). We present then the concept of *paths* in an equational proof (Section 5) and this enables us to define a measure of a goal and prove the result by induction on this measure (Section 6).

2 Preliminaries

We use standard definitions as in [1].

We will consider equations of the form $s \approx t$, where s and t are terms. Please note that throughout this paper these equations are considered to be oriented, so that $s \approx t$ is a different equation than $t \approx s$. Let E be a set of equations, and $u \approx v$ be an equation, then we write $E \models u \approx v$ (or $u =_E v$) if $u \approx v$ is true in any model containing E. We call E an equational theory, and assume that E is closed under symmetry. A goal (E-unification problem) is usually denoted by G and it is a set of equations. $E \models G$ means that $E \models e$ for all e in G.

We will be considering ground terms as ground objects that may or may not have the same syntactic form. In other words we will be concerned with the occurrences of the terms more than their values. A term may be identified by its address in a proof sequence and a position of it as a subterm in a term in the proof. Hence the equality sign between ground terms is treated in a special way. If u, v are ground terms, by $u = v$, u is understood to be an object identical with v, whereas when syntactic equality is sufficient, it will be denoted by $u == v$. Syntactic inequality will be denoted by $u \neq= v$. The difference between identity and syntactic identity is that the first involves *objects* and the second involves *names*.

[2] In [3] this rule is called *Root Rewriting*. The name *Mutate* came from [5], where it was used for E-unification in Syntactic Theories.

We can say that a variable x points to its occurrences in a term u, where each of these occurrences under some ground substitution γ, is identical with some subterm of $u\gamma$ at a position α ($x\gamma = u\gamma|_\alpha$). Different occurrences of the same variables are different objects, though they have the same syntactic form (each one is of the form $x\gamma$). In order to distinguish between different occurrences of the same variable, we will use superscript numbers, usually numbering the occurrences from left to right in order of their appearances in an equational proof. Hence $x\gamma^1$ and $x\gamma^2$ are different occurrences of x in a proof.

Sometimes we will want to state that some subterm has a form (or value) of $x\gamma$, but is not identical to $x\gamma$ (hence is not pointed to by a variable x). This will be indicated by quote marks. Hence $w[\text{“}x\gamma\text{”}]_\alpha$ is different from $w[x\gamma]_\alpha$ since in the second term $x\gamma$ actually occurs at position α, while in the first one there is only a subterm that has the value of $x\gamma$.

If γ is a ground substitution, γ_x means the restriction of this substitution to a variable x. Hence if $\gamma = [x \mapsto a, y \mapsto b, z \mapsto c]$, $\gamma_x = [x \mapsto a]$.

3 Equational Proofs

Given an equational theory E, we define an equational proof as a pair (Π, γ) such that Π is a series of ground terms and γ is a ground substitution.

Definition 1. *(Equational Proof).*
Let E be a set of equations. An equational proof of an equation $u \approx v$ is a pair (Π, γ) where $\Pi = (w_1, w_2, \ldots, w_n)$ is series of ground terms from T_{Σ_E}, called proof sequence, *such that: 1. $u\gamma = w_1$, $v\gamma = w_n$,*
2. for each pair (w_i, w_{i+1}) for $1 \leq i \leq (n-1)$, there is an equation $s \approx t \in E$ and a matcher ρ, such that there is a subterm $w_i|_\alpha$ of w_i and a subterm $w_{i+1}|_\alpha$ of w_{i+1}, and $w_i|_\alpha = s\rho$, $w_{i+1}|_\alpha = t\rho$.

We can write the equational proof as:

$$u\gamma = w_1 \approx_{[\alpha_1, s_1 \approx t_1, \rho_1]} w_2 \approx_{[\alpha_2, s_2 \approx t_2, \rho_2]} \cdots \approx_{[\alpha_{n-1}, s_{n-1} \approx t_{n-1}, \rho_{n-1}]} w_n = v\gamma$$

where $[\alpha_i, s_i \approx t_i, \rho_i]$ indicates at what position α_i is the matching subterm, which equation from E was used ($s_i \approx t_i$), and how the variables in this equation were substituted (ρ). Each w_i in the above sequence is called a term in the proof, as distinct from any proper subterms of w_i, which are not counted as terms in the proof. We will sometimes use the notation borrowed from that for arrays, and $\Pi[i]$ will mean the i'th term in Π.

Since every matcher at each step uses a renamed version of an equation from E, the domain of the matcher is disjoint from the domain of γ and the domains of matchers at all other steps in the proof, we extend γ to γ' such that: $\gamma' = \gamma \cup \rho_1 \cup \ldots \cup \rho_n$. From now on we will assume that γ is an extended version of itself.

In order to be able to identify new variables introduced by a possible application of Variable Decomposition (Figure 1), we have to extend γ even more.[3] A general extension of γ will add variables for each subterm of a term v if $\gamma_x = [x \mapsto v]$. We call these new variables *subterm variables.*

Definition 2. *(General Extension of γ).*
Let γ be a ground substitution. A general extension of γ, $ex(\gamma)$, is defined recursively as follows:
1. if $\gamma_x = [x \mapsto v]$ and $|v| = 1$ (v is a constant), then $ex(\gamma_x) = \gamma_x$,
2. if $\gamma_x = [x \mapsto f(v_1, \ldots, v_n)]$, and $n \geq 1$, then let $\gamma_{y_i} = [y_i \mapsto v_i]$, for $1 \leq i \leq n$, and $ex(\gamma_x) = \gamma_x \cup ex(\gamma_{y_1}) \cup \cdots \cup ex(\gamma_{y_n})$,
3. $ex(\gamma) = \bigcup_{x \in Dom(\gamma)} ex(\gamma_x)$

From now on we will consider γ in (Π, γ) as a general extension of itself. We have 3 kinds of variables in $Dom(\gamma)$: the *goal* variables, i.e. the variables in $Var(u \approx v)$; the *system* variables, i.e. if there is a step $\Pi[i] \approx_{[\alpha_i, s_i \approx t_i, \gamma]} \Pi[i{+}1]$ in (Π, γ), then the variables in $Var(s_i \approx t_i)$ are called *system variables*; the *subterm* variables in $\Pi[i]$, for each $\Pi[i]$ in the proof, i.e. variables that are introduced by general extension of γ. We will see that each variable occurrence starts or ends some subproof in an equational proof. In order to define this subproof, we will use a notion of *orientation* of a variable occurrence defined as follows:

Definition 3. *(Orientation of Variable Occurrences).*
Let (Π, γ) be an equational proof and $x \in Dom(\gamma)$.
1. If $x\gamma$ is a system variable occurrence in $\Pi[i] \approx_{[\alpha_i, s_i \approx t_i, \gamma]} \Pi[i + 1]$ and $x\gamma = \Pi[i]|_\alpha$ for some position α, then $x\gamma$ has left orientation. If $x\gamma = \Pi[i{+}1]|_\alpha$, then $x\gamma$ has right orientation.
2. if $x\gamma$ is a goal variable occurrence in $\Pi[1]$ ($x\gamma = \Pi[1]|_\alpha$), then $x\gamma$ has right orientation, and if $x\gamma = \Pi[n]|_\alpha$, where $\Pi[n]$ is the last term in the proof, then $x\gamma$ has left orientation.
3. if $x\gamma$ is a subterm variable, hence $x\gamma = y\gamma|_\alpha$, then $x\gamma$ has the same orientation as $y\gamma$.

3.1 Part of Equational Proof and Subproof

Now we define subproofs in an equational proof as proofs embedded at some position in *parts* of this proof.

Definition 4. *(Part of Proof for Depth α).*
Let (Π, γ) be an equational proof:

$$w_1 \approx_{[\alpha_1, s_1 \approx t_1, \gamma]} w_2 \approx_{[\alpha_2, s_2 \approx t_2, \gamma]} \cdots \approx_{[\alpha_{n-1}, s_{n-1} \approx t_{n-1}, \gamma]} w_n.$$

Let α be one of $\alpha_1, \ldots, \alpha_{n-1}$, which are the positions at which the steps in the proof are performed. A part of the proof (Π, γ) for depth α is a sequence:

[3] The following definition is similar to the definition of general extension of a substitution in [3]. It was introduced there with a similar purpose: to accommodate the Variable Decomposition rule.

$\Pi[i] \approx_{[\alpha_i, s_i \approx t_i, \gamma]} \cdots \approx_{[\alpha_{i+j-1}, s_{i+j-1} \approx t_{i+j-1}, \gamma]} \Pi[i+j]$, *such that for* $i \leq k \leq j-1$, $\alpha_k \geq \alpha$ *or* $\alpha_k || \alpha$.

Hence in a part of a proof all steps are performed at a position α, lower or at parallel positions. If $j = 0$, the part of the proof is composed of one term only. Now we define a subproof in an equational proof as a sequence of subterms of terms in a part of the original proof.

Definition 5. *(Subproof).*
Let (Π, γ) *be an equational proof. Let* $\Pi[i] \approx_{[\alpha_i, s'_1 \approx t'_1, \gamma]} \ldots \approx_{[\alpha_{i+k-1}, s'_{i+k-1} \approx t'_{i+k-1}, \gamma]}$ $\Pi[i+k]$ *be a part of the proof* (Π, γ) *for depth* α*, and let* α_n *be a such that* $\alpha \leq \alpha_n$*. Then a pair* (Σ, γ)*, where* Σ *is a sequence of terms (called* subproof sequence*):* $\Pi[i]|_{\alpha_n}, \Pi[i+1]|_{\alpha_n}, \ldots, \Pi[i+k]|_{\alpha_n}$ *is called a subproof of* (Π, γ).

In the next sections, we want to be able to use a copy of a subproof in creating new proofs. In this copy only some variables, called *internal variables*, will be renamed.

Definition 6. *(Internal/External Variables in a Subproof).*
Let (Π, γ) *be an equational proof and* $(\Sigma_{w \approx w'}, \gamma)$ *a subproof in* (Π, γ)*. If there is a step in* $(\Sigma_{w \approx w'}, \gamma)$*:* $w_i \approx_{[\alpha, s \approx t, \gamma]} w_{i+1}$*,* $y \in Var(s \approx t)$*,* y *is called an internal variable in* $(\Sigma_{w \approx w'}, \gamma)$*. If* y *has occurrences in* $(\Sigma_{w \approx w'}, \gamma)$*, but is not internal variable in this subproof, it is called an external variable in* $(\Sigma_{w \approx w'}, \gamma)$.

Definition 7. *(Renaming of a Subproof).*
Let (Π, γ) *be an equational proof and* $(\Sigma_{w \approx w'}, \gamma)$ *a subproof in* (Π, γ)*.* $(\Sigma'_{w \approx w'}, \gamma')$ *is a renaming of* $(\Sigma_{w \approx w'}, \gamma)$ *if* $(\Sigma'_{w \approx w'}, \gamma')$ *is exactly like* $(\Sigma_{w \approx w'}, \gamma)$*, with all internal variables renamed.*

Example 1. Let $E := \{ffx \approx fgx\}$ and $(\Pi, \gamma) = fgfa \approx_{[\epsilon, ffx_1 \approx fgx_1, [x_1 \mapsto fa]]} fffa \approx_{[<1>, ffx_2 \approx fgx_2, [x_2 \mapsto a]]} ffga \approx_{[\epsilon, ffx_3 \approx fgx_3, [x_3 \mapsto ga]]} fgga$.

Obviously, (Π, γ) is its own subproof. We have also one more subproof: $ffa \approx_{[\epsilon, ffx_2 \approx fgx_2, [x_2 \mapsto a]]} fga$, where $ffa = \Pi[2]|_{<1>}$. A renaming of this subproof would have the following form: $ffa \approx_{[\epsilon, ffx_4 \approx fgx_4, [x_4 \mapsto a]]} fga$, where x_4 is a new variable.

Further analysis of subproofs and their normal forms may be found in the long version of the paper [8].

3.2 Embedding a Proof into a Term and Contracting

We will use two operations on equational proofs: embedding and contracting. Embedding a proof into a term is a way to construct a proof from a given subproof.

Definition 8. *(Embedding of a Proof).*
If w *is a ground term,* (Π, γ) *is a proof of the form:*

$$w_1 \approx_{[\alpha_1, s_1 \approx t_1, \gamma]} w_2 \approx_{[\alpha_2, s_2 \approx t_2, \gamma]} \cdots \approx_{[\alpha_{n-1}, s_{n-1} \approx t_{n-1}, \gamma]} w_n$$

and there is a position β in w such that $w|_\beta == w_1$, then there is a proof (Π', γ) of the form: $w[w_1]_\beta \approx_{[\beta\alpha_1, s_1 \approx t_1, \gamma]} w[w_2]_\beta \approx_{[\beta\alpha_2, s_2 \approx t_2, \gamma]} \cdots \approx_{[\beta\alpha_{n-1}, s_{n-1} \approx t_{n-1}, \gamma]} w[w_n]_\beta$ We say that (Π', γ) is the **embedding of the proof** (Π, γ) **in the term** w.

We can attach a proof to a given equational proof (Π, γ) by embedding it into the last term of (Π, γ), if the conditions of the definition are met. If (Π, γ) is a proof such that it is composed from (Σ_1, γ_1) and (Σ_2, γ_2) by embedding (Σ_2, γ_2) into the last term of (Σ_1, γ_1), we say that (Π, γ) is a *composition* of (Σ_1, γ_1) and (Σ_2, γ_2).
A simple procedure (called **contraction**) of cutting out loops out of subproof sequences in a proof sequence, allows us to obtain a non-redundant proof from any redundant one:[4]

Definition 9. *(Non-redundant Equational Proof).*
An equational proof Π is non-redundant if there are no two terms $\Pi[i]$ and $\Pi[j]$ such that $i \neq j$ and $\Pi[i] == \Pi[j]$, and all proper subproofs of Π are non-redundant.

3.3 Associated Subproofs, Associated Terms and a Hierarchy of Variable Occurrences

In this section, for each occurrence of a variable x in $Dom(\gamma)$, we define a ground term associated with this occurrence. An associated term in the proof is a subterm which can be linked to an occurrence of $x\gamma$ by a subproof in a given equational proof. If $x \approx v$ is an equation in our goal G, and $E \models G\gamma$, then $v\gamma$ is a term associated with $x\gamma$. First, we define ground subproofs associated with each occurrence of x in an equational proof.

Definition 10. *(Subproof Associated with an Occurrence of a Variable).*
Let (Π, γ) be an equational proof, $x \in Dom(\gamma)$ and $x\gamma$ is an occurrence of x in (Π, γ).

1. *If $x\gamma$ has a left orientation and $x\gamma = \Pi[i]|_\alpha$, then there is the longest subproof*

$$\Pi[i-k]|_\alpha \approx \cdots \approx \Pi[i]|_\alpha$$

We reverse the order of the terms in this subproof:

$$\Pi[i]|_\alpha \approx \cdots \approx \Pi[i-k]|_\alpha$$

and we call this subproof **a subproof associated with this** $x\gamma$*. We say that the subproof associated with $x\gamma$ is left-oriented.*

[4] In the case of proofs in normal form, it is enough to require that there are no identical terms in the proof, to show that it is non-redundant. The definition of normal form for a proof is in the long version of the paper [8].

2. If $x\gamma$ has right orientation and $x\gamma = \Pi[i]|_\alpha$, then there is the longest subproof

$$\Pi[i+1]|_\alpha \approx \cdots \approx \Pi[i+l]|_\alpha$$

We call this subproof **a subproof associated with this** $x\gamma$ *and we say that it is right-oriented.*

Notice that if (Π, γ) is an equational proof of $u\gamma \approx v\gamma$, then the external variables in this proof are only variables in $Var(u)$ and $Var(v)$. By the definition of subproofs associated with variable occurrences, if $(\Sigma_{x\gamma\approx v}, \gamma)$ is such a subproof, external variables in this subproof have their occurrences only in $x\gamma$ (x and its subterm variables are external variables in this subproof) and v. The external variable occurrences in v have opposite orientation to that of $x\gamma$. We will sometimes indicate an orientation of an occurrence of a variable by an arrow, like in $\overrightarrow{x\gamma}$, which denotes an occurrence of x with right orientation. Similarly, if (Σ, γ) is a subproof in (Π, γ), $(\overrightarrow{\Sigma}, \gamma)$ indicates that this subproof has right orientation.

Definition 11. *(Term Associated with an Occurrence of x).*
Let (Π, γ) be an equational proof, $x \in Dom(\gamma)$ and $x\gamma$ is an occurrence of x in (Π, γ). Let a subproof $(\Sigma_{x\gamma\approx v}, \gamma)$ be a subproof associated with $x\gamma$, then we define a term associated with $x\gamma$, $ass(x\gamma)$, in the following way:

1. *if no occurrence of x appears in v, then $ass(x\gamma) = v$,*
2. *if an occurrence of x appears in v, then*
   (a) if there is a step at the root in $(\Sigma_{x\gamma\approx v}, \gamma)$, we will choose the rightmost such step: $w_i \approx_{[\epsilon, s_i \approx t_i, \gamma]} w_{i+1}$ and define $ass(x\gamma) = w_i$,
   (b) if there is no step at the root in $(\Sigma_{x\gamma\approx v}, \gamma)$, we define $ass(x\gamma) = x\gamma$.

If $x \approx w$ is an equation in a goal, where $x \notin Var(w)$, we know that there is a ground proof of $x\gamma \approx w\gamma$, and $w\gamma = ass(x\gamma)$. In this situation, we will show how to construct an equational proof of the goal with the ground substitution changed to γ', such that $\gamma'_x = [x \mapsto w\gamma]$.

There is a hierarchy among occurrences of the variables of an equational proof. In order to display it, we construct a graph G_Π with occurrences of variables in a given equational proof as nodes and arrows as follows:

1. for each variable x in $Dom(\gamma)$ and for each occurrence $x\gamma$ of this variable, if for any $y \in Dom(\gamma)$, $(\Sigma_{x\gamma\approx w[y\gamma]}, \gamma)$ is a subproof of a proof associated with $x\gamma$ and w is not empty, draw an arrow from $x\gamma$ to $y\gamma$;

2. for each variable x in $Dom(\gamma)$ and for each occurrence $x\gamma$ of this variable, if for any $y \in Dom(\gamma)$, $(\Sigma_{x\gamma\approx y\gamma}, \gamma)$ is a subproof of a proof associated with $x\gamma$:

2.1 if $(\Sigma_{y\gamma\approx x\gamma}, \gamma)$ is a subproof of a proof associated with $y\gamma$, then nondeterministically decide the direction of an arrow between $x\gamma$ and $y\gamma$;

2.2 if $(\Sigma_{y\gamma\approx x\gamma}, \gamma)$ is not a subproof associated with $y\gamma$, then draw an arrow from $x\gamma$ to $y\gamma$.

The graph G_Π, for an equational proof helps us to recognize/decide the parent/child relation. If $x\gamma$ is a node in G_Π and there is an arrow $x\gamma \to y\gamma$, then $x\gamma$ is called a parent of $y\gamma$ and $y\gamma$ is a child of $x\gamma$.

This relation is in some cases determined by the structure of the proof (we cannot discover new variables in the transformation of the goal before solving/eliminating some other variables first), or it is decided by the selection rule and orientation of an equation of the form $x \approx y$. The *maximal* nodes in the graph are just those occurrences of variables that are discovered in the goal and may be selected for eager Variable Elimination. A set of maximal nodes in G_{Π}, M, is the set containing all nodes which have no parents in G_{Π}. We will see that if $x \approx v$ is an equation in a goal G, and $E \models G\gamma$, then there is a subproof $(\Sigma_{x\gamma\approx v\gamma}, \gamma)$ in an equational proof (Π, γ) of $G\gamma$ and $x\gamma$ is a maximal node in the graph G_{Π} for the proof.

4 Solving Variables in an Equational Proof

The following construction explains what happens with an equational proof of a goal, if an equation of the type $x \approx t$ is selected for eager Variable Elimination. Notice that in this construction we declare which variables in $Dom(\gamma)$ are solved or unsolved. In the justification of the completeness of the inference system with eager Variable Elimination we start with the equational proof of an instance of a goal with all variables unsolved. Variable Elimination reflects solving variables in a ground equational proof.

Let (Π, γ) be an equational proof with the proof sequence:

$$\Pi = (w_1 \approx_{[\alpha_1, s_1 \approx t_1, \gamma]} w_2 \approx_{[\alpha_2, s_2 \approx t_2, \gamma]} \dots \approx_{[\alpha_{n-1}, s_{n-1} \approx t_{n-1}, \gamma]} w_n)$$

and γ be an extended ground substitution.

Let $U = \{x_1, \dots, x_n\}$ be a set of variables called "unsolved" in (Π, γ), G_{Π} be the graph for (Π, γ) constructed only with respect to unsolved variables (hence we treat all other variables as non-existent in (Π, γ)).

Let $x \in U$ and $x\gamma$ be a maximal node in G_{Π} and let $ass(x\gamma) = v$.

There is a subproof $(\Sigma_{x\gamma\approx v}, \gamma)$ in (Π, γ), let $(\Sigma'_{x\gamma\approx v}, \gamma')$ be a renaming of this subproof.[5]

If x has no occurrences in v, create a new proof (Π^*, γ^*) that is exactly as (Π, γ) with the proof sequence modified in the following way:

1. **Extension**
   Whenever $x\gamma = w_i|_{\alpha}$ and hence $w_i = w_i[x\gamma]$, and
   (a) *$x\gamma$ has right orientation, replace w_i (the i'th step in (Π, γ)), by the sequence of steps:*

$$w_i[v]_{\alpha} \approx (\Sigma'_{v\approx``x\gamma"}) \approx w_i[``x\gamma"]_{\alpha}$$

   where $(\Sigma'_{v\approx``x\gamma"})$ means a renaming of $(\Sigma_{``x\gamma"\approx v}, \gamma)$ reversed and embedded in w_i at position α leftwards. Note that the renamings of internal occurrences of variables and new occurrences of external variables in the renaming of $(\Sigma_{``x\gamma"\approx v}, \gamma)$ have reversed orientation in the new proof.

[5] If x has no occurrences in v, $(\Sigma_{x\gamma\approx v}, \gamma)$ is a subproof associated with $x\gamma$.

(b) *$x\gamma$ has left orientation, replace w_i (the i'th step in (Π, γ)) by the sequence of steps:*

$$w_i[\text{“}x\gamma\text{”}]_\alpha \approx (\Sigma'_{\text{“}x\gamma\text{”}\approx v}) \approx w_i[v]_\alpha$$

where $(\Sigma'_{\text{“}x\gamma\text{”}\approx v})$ means a renaming of $(\Sigma_{\text{“}x\gamma\text{”}\approx v}, \gamma)$ embedded in w_i at position α rightwards. The renamings of internal occurrences of variables and new occurrences of external variables in $(\Sigma'_{\text{“}x\gamma\text{”}\approx v})$ preserve their orientation in the new proof.

2. **Standard Contraction**
For each occurrence of an unsolved variable y in (Π, γ), if $(\Sigma_{y\gamma\approx s}, \gamma)$ is a proper associated subproof of this occurrence in (Π, γ) and there is a subproof sequence: $\Sigma_{s\approx\text{“}y\gamma\text{”}} \Sigma_{\text{“}y\gamma\text{”}\approx s}$ in the proof sequence Π^* after extension, contract the subproof sequence to a one-element sequence, s;

The substitution γ^* is defined as follows:
$\gamma^*_x = [x \mapsto v]$,
if $y\gamma|_\alpha = x\gamma$, and $y \notin U$, then $\gamma^*_y = [y \mapsto y\gamma[x\gamma^*]_\alpha]$,
if $z \notin Dom(\gamma)$, z is a renaming of a variable $z' \in Dom(\gamma)$, that appeared in some $(\Sigma'_{x\gamma\approx v}, \gamma')$, then $\gamma^*_z = [z \mapsto z'\gamma]$,
for any other variable, $\gamma^* = \gamma$;
If x has occurrences in v, then $(\Pi^*, \gamma^*) = (\Pi, \gamma)$.

Mark Variables
Mark variable x **solved** *in (Π^*, γ^*). If x has no occurrences in v, mark also all subterm variables of x as* **solved** *New variables in $Dom(\gamma^*)$, which did not appear in $Dom(\gamma)$ are marked as unsolved.*

If a proof (Π^*, γ^*) is obtained from (Π, γ) in this way, then we say that (Π^*, γ^*) is generated from (Π, γ) by substitution $[x \mapsto v]$, written $(\Pi, \gamma) \overset{[x\mapsto v]}{\rightarrow} (\Pi^*, \gamma^*)$. As a corollary to this construction we notice that:

Corollary 1. *If $(\Pi, \gamma) \overset{[x\mapsto v]}{\rightarrow} (\Pi', \gamma')$ and $y \in Dom(\gamma')$, then for each occurrence $y\gamma'$ in (Π', γ'), either*

1. *$y \in Dom(\gamma)$ and $y\gamma'$ is an occurrence of this variable identical with an occurrence in (Π, γ), ($y\gamma'$ is in the part of (Π', γ') not affected by extension and contraction), or*
2. *$y \in Dom(\gamma)$ and $y\gamma'$ is a new occurrence of y, introduced in the effect of extending (Π, γ) with $(\Sigma_{x\gamma\approx v}, \gamma)$, (there was an occurrence $y\gamma^k$ of an external variable y in $(\Sigma_{x\gamma\approx v}, \gamma)$ which generated new occurrences in all places the copy of this subproof was used and not contracted), or*
3. *$y \notin Dom(\gamma)$, (y is a new variable) then $y\gamma'$ may be identified as a renamed version of a variable $y' \in Dom(\gamma)$, where y' was an inner variable in $(\Sigma_{x\gamma\approx v}, \gamma)$.*

Example 2. Let an equational proof be:

$$f(a, g(b, b)) \approx_{[<1>, a\approx b, []]} f(b, g(b, b)) \approx_{[\epsilon, f(x,g(x,x))\approx c, [x\mapsto b]]} c$$

Then the subproof associated with $\overleftarrow{x\gamma}^1$ is $b \approx a$. Notice the left orientation of all occurrences of x in this case. We want to solve x in the proof with $x \mapsto a$. Hence we will use $b \approx a$ for the extension at each occurrence of x.

$$\begin{aligned} f(a, g(b, b)) &\approx_{[<1>, a\approx b, []]} f(b, g(b, b)) \approx_{[<1>, b\approx a, []]} f(a, g(b, b)) \\ &\approx_{[<2>, b\approx a, []]} f(a, g(a, b)) \approx_{[<3>, b\approx a, []]} f(a, g(a, a)) \\ &\approx_{[\epsilon, f(x, g(x,x))\approx c, [x\mapsto a]]} c \end{aligned}$$

Standard contraction will shorten the proof to:

$$\begin{aligned} f(a, g(b, b)) &\approx_{[<2>, b\approx a, []]} f(a, g(a, b)) \approx_{[<3>, b\approx a, []]} f(a, g(a, a)) \\ &\approx_{[\epsilon, f(x, g(x,x))\approx c, [x\mapsto a]]} c \end{aligned}$$

Notice that we have a new assignment for x, but now we will treat x as solved.

5 Paths in Equational Proof

A concept of path is a generalization of an associated subproof for an occurrence of a variable. A path is a subproof starting with some variable occurrence, constructed in such a way that it reflects the form of an associated subproof for this variable occurrence assuming that all other variables *involved* in the path were solved first. In order to restrict the definition of a path in a proof (Π, γ), we have to take into consideration solved and unsolved occurrences of variables in $Dom(\gamma)$. We have to remember where the solved variables had their occurrences at the time they were being solved.

Since in this section we will deal with compositions of subproofs, in order to simplify notation, we will identify a subproof with its subproof sequence.

Definition 12. *(Path Starting with a Variable Occurrence and Variable Occurrence Involved in a Subproof).*
Let (Π, γ) be an equational proof, U a set of unsolved variables in $Dom(\gamma)$, $x \in U$ and $x\gamma$ a given variable occurrence in (Π, γ). A path in (Π, γ) starting with $x\gamma$ is a composition of subproofs, $\Sigma_1 \dots \Sigma_n$, defined in a recursive way:

1. *if y is a solved variable and $\Sigma_{v_1 \approx \text{``}y\gamma'^i\text{''}} \Sigma_{\text{``}y\gamma'^k\text{''} \approx v_2}$ is a subproof in (Π, γ), then $y{\gamma'}^k$ is an occurrence of a solved variable involved in this subproof;*
2. *if $\Sigma_{x\gamma \approx v}$ is an associated subproof for $x\gamma$, $\Sigma_{x\gamma\approx v}$ is a path starting with $x\gamma$ and $x\gamma$ is involved in this path;*
3. *if $y\gamma$ is a parent of $x\gamma$, then $\Sigma_{y\gamma \approx w[x\gamma]}$ is a path starting with $y\gamma$ and $y\gamma$ is involved in this path.*
4. *(a) if $\Sigma_1, ..., \Sigma_n$ is a path in (Π, γ) starting with $x_1\gamma$ and $\Sigma_n = \Sigma_{x_n\gamma \approx v[x_{n+1}\gamma^k]}$, and if x_{n+1} is an external variable in $\Sigma_{x_n\gamma \approx v[x_{n+1}\gamma^k]}$ and $\Sigma'_1, \dots, \Sigma'_m$ is a path in (Π, γ) starting with $x_{n+1}\gamma^i$, and if no variable occurrence that is involved in the first part is involved in the second, and vice versa then the composition $\Sigma_1 \dots \Sigma_n \Sigma'_1 \dots \Sigma'_m$ is also a path in (Π, γ) starting with $x_1\gamma$ and all variables involved in the first and second path are involved in the new path;*

(b) if $\Sigma_1, \ldots, \Sigma_n$ is a path in (Π, γ) starting with $x_1\gamma$ and $\Sigma_n = \Sigma_{x_n\gamma \approx y\gamma|_\alpha}$, and $\Sigma_{y\gamma^k|_\alpha \approx s}$ is a subproof in (Π, γ) and if no variable occurrence that is involved in $\Sigma_1, \ldots, \Sigma_n$ is involved in $\Sigma_{y\gamma^k|_\alpha \approx s}$, and vice versa then $\Sigma_1, \ldots, \Sigma_n, \Sigma_{y\gamma^k|_\alpha \approx s}$ is also a path in (Π, γ) starting with $x_1\gamma$ and all variable occurrences involved in $\Sigma_1, \ldots, \Sigma_n$, in $\Sigma_{y\gamma^k|_\alpha \approx s}$ and $y\gamma^k$ are involved in the new path;

5. *if Π_l is a path or a subproof in a path, then any composition $\overleftarrow{\Pi_l}\overrightarrow{\Pi_l}$ in the path is contracted in an obvious way to one element (part of) path.*

Notice that an occurrence of a variable is involved in a path if it is a *beginning* of a subpath. An occurrence of a variable that appears at the end of a path is not necessarily involved in this path. An occurrence of a variable that is involved in a path is an entry point to some subpath of this path.

Example 3. For example, let our goal be: $G = \{x \approx a, z \approx hx, z \approx c\}$ and an equational theory: $E = \{b \approx a, b \approx fga, hfy \approx c\}$, then the proof (Π, γ) may be:

$$\begin{array}{ccccccc} x\gamma^1 & & x\gamma^2 & & & & \\ \downarrow & & \downarrow & & & & \\ \{\, b & \approx_{[\epsilon, b\approx a, []]} a, & hb, & hb \approx_{[<1>, b\approx fga, []]} & hfga & \approx_{[\epsilon, hfy\approx c, [y\mapsto ga]]} & c\,\} \\ & & \uparrow & \uparrow & & & \\ & & z\gamma^1 & z\gamma^2 & & & \end{array}$$

1. An example of a path starting with $x\gamma^2$ would be: $\Sigma_{x\gamma^2 \approx z\gamma^1|_{<1>}} \Sigma_{z\gamma^2|_{<1>} \approx fy\gamma}$.
2. An example of a path starting with $z\gamma^1$ is: $\Sigma_{z\gamma^1 \approx h(x\gamma^2)} \Sigma_{x\gamma^1 \approx a}$.

We will prove that if $(\Pi, \gamma) \overset{[x\mapsto v]}{\rightarrow} (\Pi', \gamma')$, for an unsolved variable x in $Dom(\gamma)$, then each path in (Π', γ') starting with an unsolved variable in (Π', γ') is identical to a path in (Π, γ) (up to renaming). Hence any new paths will be renamings of the original ones. In order to show that the process of solving variables in (Π, γ) will terminate, we will use a multiset of lengths of paths as a measure, and show that it is decreasing.

Lemma 1. *Let (Π, γ) be an equational proof, $U \subset Dom(\gamma)$ be a set of unsolved variables in (Π, γ), $(\Pi, \gamma) \overset{[x\mapsto v]}{\rightarrow} (\Pi', \gamma')$, and U' be a set of unsolved variables in (Π', γ').*

Each path in (Π', γ') starting with a variable occurrence of a variable in U' is identical (up to renaming of some variables) to a path in (Π, γ) starting with a variable occurrence of a variable in U.

If there are many paths in (Π', γ'), which are renamings of one and the same path in (Π, γ), then they are strictly shorter than a path in (Π, γ), starting with a variable occurrence of a variable which is solved in (Π', γ').

Proof. The proof of this lemma is based on the fact that each path starting with an occurrence of an unsolved variable in (Π, γ) is finite. Hence we can use induction on the lengths of paths.

Let $(\Pi, \gamma) \stackrel{[x \mapsto v]}{\to} (\Pi', \gamma')$, where $\Sigma_{x\gamma^i \approx v}$ was used in construction of (Π', γ'), and $\Sigma_1 \dots \Sigma_n$ is a path in (Π', γ') starting with $y_1\gamma$. We can assume that x does not occur in v, because otherwise $ass(x\gamma) = x\gamma$ and then (Π', γ') is identical to (Π, γ) with the only difference that x is solved and does not appear in U'.

We have to consider different cases connected with the possible ways the path $\Sigma_1 \dots \Sigma_n$ was constructed in (Π', γ'). The full proof and examples are in the long version of the paper [8].

Here, as an example, we will see the first of the cases considered there, namely when $\Sigma_1 \dots \Sigma_n$ is a path starting with $y_1\gamma$ by Definition 12.2. Hence $\Sigma_1 \dots \Sigma_n$ is a subproof associated with $y_1\gamma$ in (Π', γ'). The only case, when such a subproof was not a path in (Π, γ) would be if the composition of shorter paths was prevented by the condition that an occurrence of a variable may be involved only once in a path. Hence $\Sigma_1 \dots \Sigma_n$ would have to be composed from two shorter paths: $\Sigma_{y\gamma' \approx t[``x\gamma^k"]}$ and $\Sigma_{``x\gamma^i" \approx v}$ and in both of them an occurrence of a variable z is involved. Since $\Sigma_{x\gamma^i \approx v}$ is an associated proof for $x\gamma^i$, the occurrence of z would have to be an occurrence of a solved variable. The both paths would have to be of the following forms: $\Sigma_{y_1'\gamma \approx s[``z\gamma^k"]} \Sigma_{``z\gamma^i" \approx t[x\gamma^i]}$, $\Sigma_{x\gamma^i \approx v[``z\gamma^{*k}"]} \Sigma_{``z\gamma^{*i}"} \approx t[x\gamma]$. Hence v contains occurrence of x contrary to our assumption.

Corollary 2. *Let (Π, γ) be an equational proof, $U \subset Dom(\gamma)$ a set of unsolved variables in (Π, γ), The process of solving (Π, γ) will terminate.*

Proof. If $(\Pi, \gamma) \stackrel{[x \mapsto v]}{\to} (\Pi', \gamma')$, and U' a set of unsolved variables in (Π', γ'), the multiset of lengths of paths in (Π', γ') is smaller than the multiset of lengths of paths in (Π, γ).

6 Result

We prove completeness of the inference rules presented in Figure 1.

Namely, we prove that in any equational theory E, a given goal G such that $E \models G\sigma$, may be transformed by applications of rules in Figure 1 applied to equations which are not *solved*, into a *solved form* with which we can define an E-unifier more general than σ. The solved form of an equation and of a goal is defined in the following way.

Definition 13. *(Solved Equation and Solved Goal).*
Let G be a set of equations. An equation $x \approx t \in G$ is in a solved form*, if x is a variable, $x \notin Var(t)$ and $x \notin Var(G \backslash \{x \approx t\})$.*

G is in a solved form if all equations in G are in solved form.

If G is in the solved form, then we define a substitution $\theta_G = [x_1 \mapsto t_1, \dots, x_n \mapsto t_n]$. Obviously, θ_G is the most general unifier of G.

If G is a set of goal equations, an inference performed on G with one of the rules of Figure 1 is denoted by $G \to G'$, where G' is the result of this inference. The transitive, reflexive closure of $\to$ is written as $\stackrel{*}{\to}$.

In order to prove completeness, we will need the measure of a goal G, of which we will show that it may be decreased by application of an inference rule if G is E-unifiable and not in solved form.

Definition 14. *(Measure for an Equational Proof).*
Let (Π, γ) be an equational proof and $U \subset Dom(\gamma)$ be a set of unsolved variables in (Π, γ). The measure $M(\Pi, \gamma)$ is a multiset of the lengths of paths starting with occurrences of variables in U.

Definition 15. *(Measure of a Goal).*
Let E be an equational theory, and G, an unsolved part of a goal G', such that there is a ground substitution γ, for which $E \models G'\gamma$ and hence there is an equational proof (Π', γ') of $G'\gamma$ and its subproof, (Π, γ), which is a proof of $G\gamma$, and all variables in $Var(G)$ are unsolved in (Π', γ').

The measure of G' with respect to (Π', γ') is a 4-tuple (m, n, o, p), where $m = M(\Pi, \gamma)$, n is the length of Π, o is the size of terms in $G\gamma$, p is the number of equations in G, of the form $t \approx x$, where x is a variable and t is not a variable.

Measures for different goals are compared with respect to lexicographic order.

Theorem 1. *Let E be a set of equations, such that $E \models G\gamma$ for some ground substitution γ. Then there is H a set of equations in the solved form, such that $G \overset{*}{\to} H$ and $\theta_H[Var(G)] \leq_E \gamma$.*

Proof. If G is already in the solved form, then $\theta_G \leq_E \gamma$.

If G is not in solved form, then there is an unsolved part of G, G', such that $u \approx v \in G'$, if $u \approx v$ is not in solved form. Assume that $u \approx v$ was selected for an inference. If $E \models G\gamma$, there must be an equational proof (Π, γ) of $G'\gamma$. We will call it an actual proof of $G\gamma$. If $u\gamma \approx v\gamma \in G\gamma$, then there must be a subproof in (Π, γ), of this ground equation and $u\gamma$, $v\gamma$ are the extreme terms in this subproof. We can also assume that all solved variables in G are solved in (Π, γ) and all unsolved variables in G are unsolved in (Π, γ). Hence there is a graph G_Π for all unsolved variables in (Π, γ). As we have seen, there are sometimes choices in constructing G_Π. The choices reflect the selection function, but in any case, we can always choose such G_Π that if $x \approx v$ is selected for an inference, $x\gamma$ is a maximal node in G_Π.

The full proof which uses induction on the measure of a goal, is in the long version of the paper [8]. Here we will see only the case for eager Variable Elimination.

Assume that $x \approx v$ was selected for an inference and $x \notin Var(v)$. Then $E \models x\gamma \approx v\gamma$ and there is a subproof $(\Sigma_{x\gamma \approx v\gamma}, \gamma)$ in the proof (Π, γ) such that $x\gamma$ and $v\gamma$ are extreme terms of $(\Sigma_{x\gamma \approx v\gamma}, \gamma)$. If x is unsolved in the goal G, x is also unsolved in (Π, γ). Hence we know that $v\gamma = ass(x)$ and $(\Pi, \gamma) \overset{[x \mapsto v\gamma]}{\to} (\Pi', \gamma')$. $M(\Pi, \gamma) > M(\Pi', \gamma')$.

Since $E \models G\gamma$, also $E \models G\gamma'$ and (Π', γ') is the proof of $G\gamma'$. We change the actual equational proof to (Π', γ') and take it as the basis of completeness argument of further inferences. Since $x\gamma' = v\gamma'$, $E \models G[x \mapsto v]\gamma'$.

Let (m, n, o, p) be the measure of the goal before Variable Elimination and (m', n', o', p') after Variable Elimination. $m' < m$ after Variable Elimination, hence we can use induction hypothesis for the new goal. Notice also that after Variable Elimination, for each $u' \approx v'$ in G' there is a subproof in (Π', γ') such that $u'\gamma'$ and $v'\gamma'$ are the extreme terms in this subproof. If $u' \approx v'[x]$ was in G', then after Variable Elimination, $u' \approx v'[v]$ in G' and obviously (because of extension) there is a subproof $(\Sigma_{u'\gamma' \approx v'[v]\gamma'}, \gamma')$ in (Π', γ').

7 Conclusion

E-unification procedures are inherently non-deterministic, because there are usually many ways to apply inferences to goal equations and many possibilities of solving a goal. It means that a search space for a solution may be very extensive. Any restrictions of this non-determinism that we may justify are therefore welcome as restrictions of this search space. Eager Variable Elimination means that the rule should be applied whenever an equation $x \approx v$ is selected and x does not appear in v. In this case, we would not try to apply other rules to this equation. On the other hand, we may see that the ground equational proof of an instance of a goal, may be made longer by Variable Elimination. This means that we will have to do more Mutate inferences in order to reach solution. One can think about some *memoization* techniques to detect and reduce such possible overhead.

We think that the proof of completeness of eager Variable Elimination opens some possibilities of finding new classes of equational theories defined syntactically, for which E-unification problem may be proved solvable and tractable.

References

1. F. Baader and T. Nipkow. *Term Rewriting and All That.* Cambridge, 1998.
2. J. Gallier and W. Snyder. A general complete E-unification procedure. In *RTA 2*, ed. P. Lescanne, LNCS Vol. 256, 216-227, 1987.
3. J. Gallier and W. Snyder. Complete sets of transformations for general E-unification. In *TCS*, Vol. 67, 203-260, 1989.
4. S, Hölldobler. Foundations of Equational Logic Programming. *Lecture Notes in Artificial Intelligence*, Vol. 353, Springer, Berlin, 1989.
5. C. Kirchner and H. Kirchner. *Rewriting, Solving, Proving.* `http://www.loria.fr/~ckirchne/` , 2000.
6. A. Martelli, C. Moiso and G. F. Rossi. Lazy Unification Algorithms for Canonical Rewrite Systems. In *Resolution of Equations in Algebraic Structures*, eds. H. Aït-Kaci and M. Nivat, Vol. II of Rewriting Techniques, 258-282, Academic Press, 1989.
7. A. Martelli, G. F. Rossi and C. Moiso. An Algorithm for Unification in Equational Theories. In *Proc. 1986 Symposium on Logic Programming*, 180-186, 1986.
8. B. Morawska. Completeness of E-unification with eager Variable Elimination. LTCS-Report (03-03), See http://lat.inf.tu-dresden.de/research/reports.html.