NATURAL TRANSFORMATIONS AS REWRITE RULES
AND MONAD COMPOSITION

DEXTER KOZEN

Computer Science Department, Cornell University, Ithaca, NY 14853-7501, USA
e-mail address: kozen@cs.cornell.edu

In honor of Jiří Adámek on the Occasion of his Seventieth Birthday.

Abstract. Eklund et al. (2002) present a graphical technique aimed at simplifying the verification of various category-theoretic constructions, notably the composition of monads. In this note we take a different approach involving string rewriting. We show that a given tuple \((T, \mu, \eta)\) is a monad if and only if \(T\) is a terminal object in a certain category of strings and rewrite rules, and that this fact can be established by proving confluence of the rewrite system. We illustrate the technique on the monad composition problem. We also give a characterization of adjunctions in terms of rewrite categories.

1. INTRODUCTION

As common constructions in the theory of data types and programming language semantics become better understood, there is a natural tendency toward generality. One desires to isolate common underlying principles, to unify related notions in a common framework, and to provide powerful abstract tools for reasoning and understanding.

A good example of one successful such enterprise is the use of monads in functional and logic programming [3, 7, 11, 17]. Monads provide a clean way to combine modules or extend functionality of programming languages or data structures with new features such as continuations, state, and concurrency [6, 8, 11, 17]. They have been applied to parsing and type checking [16], semantics of nondeterministic and probabilistic computation [11–13], and unification in logic programming [14].

Unfortunately, greater abstraction is often accompanied by reduced accessibility. Many abstract constructions, although well motivated by applications, may at times be difficult to navigate when presented in more abstract form. In particular, reasoning about the basic properties of monads—such as monad composition, the construction that underlies many of the applications above—relies on the combinatorial manipulation of functors and natural transformations. Verification often requires a complicated process of arrow chasing in large diagrams. Specialized verification tasks such as the following example from [4] are not uncommon.


Key words and phrases: Monad; monad composition; rewriting.
There are several general accounts of monad composition in the literature \[3, 7, 10\]. In particular, Eklund et al. \[7\] make note of the difficulty of monad composition proofs and present a graphical technique aimed at simplifying the process. Their technique provides a pictorial representation of various constructions.

In this note we take a different approach involving string rewriting. We show that a given tuple \((T, \mu, \eta)\) is a monad if and only if \(T\) is a terminal object in a certain category whose objects are strings and whose morphisms are rewrite rules, and the fact that \(T\) is terminal can be established by proving confluence of the string rewriting system. We illustrate the technique on the monad composition problem studied by Eklund et al. We also give a characterization of adjoint functors in terms of rewrite categories.

2. Preliminaries

Let \(C, D\) be categories. Recall that a natural transformation \(\varphi : F \to G\) between functors \(F, G : C \to D\) is a collection of morphisms \(\varphi_A : FA \to GA\) of \(D\), one for each object \(A\) of \(C\), such that for any morphism \(h : A \to B\) of \(C\), the following diagram commutes:

\[
\begin{array}{ccc}
FA & \xrightarrow{Fh} & FB \\
\varphi_A \downarrow & & \downarrow \varphi_B \\
GA & \xrightarrow{Gh} & GB
\end{array}
\]  

(2.1)

Categories, functors, and natural transformations form a 2-category \(\text{Cat}\) in which the 0-cells (objects) are the categories, the 1-cells (morphisms) are the functors, and the 2-cells (morphisms of morphisms) are the natural transformations.

In a general 2-category, composition of 1-cells and composition of 2-cells are called horizontal and vertical composition, respectively, and are often denoted \(\circ_0\) and \(\circ_1\), respectively; but we will usually write \(\circ\) for \(\circ_1\) and omit the \(\circ_0\) altogether. Horizontal composition also acts on 2-cells and satisfies the property: If \(F_1, F_2, F_3 : C \to D\) and \(G_1, G_2, G_3 : D \to E\) are 1-cells and \(\varphi : F_1 \to F_2, \varphi' : F_2 \to F_3, \psi : G_1 \to G_2, \text{ and } \psi' : G_2 \to G_3\) are 2-cells, then

\[
(\psi' \circ \psi)(\varphi' \circ \varphi) = (\psi' \varphi')(\varphi \psi).
\]  

(2.2)

A special case of the action of horizontal composition on 2-cells is the following. If \(S : D \to E, F, G : C \to D\), and \(T : B \to C\) are 1-cells and \(\varphi : F \to G\) is a 2-cell, there is a 2-cell \(S\varphi T : SFT \to SGT\) obtained from the horizontal composition of \(\varphi\) with the identities \(\text{id}_S\) on the left and \(\text{id}_T\) on the right. It is helpful to think of this operation as a kind of scalar
multiplication. It satisfies the following properties:

\[ I \varphi = \varphi = \varphi I \]
\[ (SS')\varphi = S(S'\varphi) \quad \varphi(TT') = (\varphi T)T' \]
\[ S(\varphi \circ \psi) = S\varphi \circ S\psi \quad (\varphi \circ \psi)T = \varphi T \circ \psi T. \]

Thus we can write \( SS'\varphi \) and \( \varphi TT' \) without ambiguity.

**Example 2.1.** In the 2-category \( \text{Cat} \), if \( S \) and \( T \) are functors and \( \varphi \) is a natural transformation with components \( \varphi_A \), then \( S\varphi T \) is the natural transformation with components \( (S\varphi T)_A = S(\varphi_T A) \).

It follows from (2.3) that the commutativity of diagrams is preserved under scalar multiplication; that is, if the left-hand diagram in (2.4) commutes, then so does the right:

\[ \begin{array}{ccc}
F & \varphi & G \\
\psi & & \\
X & \sigma & Y \\
\end{array} \quad \begin{array}{ccc}
SFT & \xrightarrow{\varphi T} & SGT \\
S\psi T & \xrightarrow{\psi T} & S\tau T \end{array} \]

From these observations, one can begin to see the motivation for viewing natural transformations as rewrite rules: a rewrite rule \( \varphi : F \to G \) can be applied in the context of a string \( S\varphi T : SFT \to SGT \). For this to make sense, though, it had better be the case that the diamond property holds for non-overlapping redexes; that is, it must be possible to apply two rewrite rules with non-overlapping redexes in either order. We will see later (Lemma 3.5) that this is indeed the case, and in fact holds in all 2-categories as a consequence of (2.2).

### 3. Rewrite Categories

A **rewrite category** is like a string rewrite system, except the semantics of the derivations are taken into account in the definitions of confluence and local confluence.

Formally, a rewrite category is defined in terms of a finitely presented 2-category \((O, F, R, E)\), where \( O \), \( F \), and \( R \) are finite sets of 0-cells, 1-cells, and 2-cells, respectively, and \( E \) is a finite set of well-typed equations between 2-cell-valued expressions over \( F \) and \( R \) generating an equational theory on 2-cells. The 1-cells do not satisfy any equations except those imposed by the axioms of category theory, thus form the free category \((\text{free typed monoid})\) freely generated by \( F \). Each 1-cell corresponds to a well-typed string \( F_1 \cdots F_n \), where well-typed means \( \text{dom} F_i = \text{cod} F_{i+1} \), \( 1 \leq i \leq n-1 \). We denote by \( F^* \) the set of all well-typed strings over \( F \). The elements of \( R \) are called rewrite rules and generate the set of all 2-cells under horizontal \((\circ_0)\) and vertical \((\circ_1)\) composition.

The rewrite category itself is a substructure of the category of 1- and 2-cells of this finitely presented 2-category. It is specified by a tuple \((A; R')\), where \( R' \subseteq R \) and \( A \) is any subset of \( F^* \) closed under the action of \( R' \).

Rewrite categories provide a syntax that can be interpreted in the 2-category \( \text{Cat} \) of categories, functors, and natural transformations. A well-typed string \( F_1 \cdots F_n \) is interpreted as a functor \( \text{dom} F_n \to \text{cod} F_1 \), where concatenation of strings is interpreted as composition of functors. The rewrite rules describe the action of natural transformations on these strings.
The equations $\mathcal{E}$ generating the equational theory of the 2-category typically represent an equational axiomatization of the construct under study. However, from the perspective of string rewriting, they serve a different purpose: they represent local confluence conditions that can be used to establish global confluence in the rewrite category, leading to the existence of terminal objects.

**Example 3.1.** Recall that a monad on a category $C$ is a triple $(T, \mu, \eta)$, where $T$ is an endofunctor on $C$ and $\mu : T^2 \to T$ and $\eta : I \to T$ are natural transformations such that $\mu \circ \mu T = \mu \circ T \mu$ and $\mu \circ \eta T = \mu \circ T \eta = \text{id}$; that is, the diagrams

$$
\begin{align*}
T^3 & \xrightarrow{T \mu} T^2 \\
\mu T & \xrightarrow{\mu} T
\end{align*}
\quad
\begin{align*}
T & \xrightarrow{\eta} T^2 \\
\eta T & \xrightarrow{id} \mu T \xrightarrow{\mu} T
\end{align*}
$$

(3.1)

commute. A typical example is the powerset monad $(P, \mu^P, \eta^P)$ on $\text{Set}$, where $PA$ is the powerset of $A$, $\eta^P_A(x) = \{x\}$, and $\mu^P_A(C) = \bigcup C$.

Monads are characterized by a rewrite category $(T^*; \eta, \mu)$, where $T^*$ is the free monoid on one generator $T$ with rewrite rules $\eta : I \to T$ and $\mu : T^2 \to T$. The underlying 2-category has one 0-cell. The equational theory on 2-cells is given by the defining equations for monads (3.1).

The properties (3.1) completely axiomatize the equational theory of monads, but more importantly from the point of view of rewriting, they specify local confluence properties in the case of overlapping redexes. For example, the string $T^3$ contains two overlapping redexes for the rule $\mu$. The left-hand diagram of (3.1) specifies that this configuration satisfies the diamond property.

We will show later (Theorem 4.2) that $T$ is a terminal object in the rewrite category, and that this fact is equivalent to the axiomatization (3.1).

**Example 3.2.** The composition of monads is characterized by a rewrite category $(\{P, T\}^*; \eta^P, \mu^P, \eta^T, \mu^T, \theta)$, where $\{P, T\}^*$ is the free monoid on two generators $P, T$. As in Example 3.1, the underlying 2-category has one 0-cell. The equational theory is given by the axiomatization of two monads on endofunctors $P$ and $T$, respectively, and a distributive law $\theta : TP \to PT$ connecting them. We will give more details in §4.2. Again, the axioms can be viewed as local confluence properties for overlapping redexes in the rewrite category.

**Example 3.3.** Adjunctions are characterized by a rewrite category $(F(G)*) \cup G(FG)*; \eta, \varepsilon)$. In this example, the underlying 2-category has two 0-cells $C, D$ representing two categories, 1-cells $F : C \to D$ and $G : D \to C$ representing left and right adjoint functors, respectively, and 2-cells $\eta : I \to GF$ and $\varepsilon : FG \to I$ representing the unit and counit of the adjunction, respectively.

### 3.1. Local Confluence and Normal Forms

Given a 1-cell $SXT$ and a rule $\varphi : X \to Y$ of a rewrite category, the scalar multiple $S\varphi T : SXT \to SYT$ can be viewed as rewriting the indicated occurrence of $X$ to $Y$ in the context of the string $SXT$. This is called a reduction. The redex of the reduction $S\varphi T$ is the substring $X$ of $SXT$. A sequence of reductions from
Y to Z is called a derivation. We write \( \pi : Y \overset{*}{\to} Z \) if \( \pi \) is such a derivation. A string is in normal form if no rules apply.

The rules \( \mathcal{R} \) of the rewrite category each have a type \( X \to Y \) for various \( X, Y \in \mathcal{F}^* \), where \( \text{dom} \ X = \text{dom} \ Y \) and \( \text{cod} \ X = \text{cod} \ Y \). Applying these rules to strings in \( \mathcal{F}^* \), one can generate diagrams in the rewrite category. We say that the system \( \mathcal{R} \)

- is confluent if for any two derivations \( X \overset{*}{\to} U \) and \( X \overset{*}{\to} V \), there is a word \( W \) and derivations \( U \overset{*}{\to} W \) and \( V \overset{*}{\to} W \) such that the resulting diagram commutes;
- is locally confluent if for any two single reductions \( X \to U \) and \( X \to V \), there is a word \( W \) and derivations \( U \overset{*}{\to} W \) and \( V \overset{*}{\to} W \) such that the resulting diagram commutes; and
- has the diamond property if for any two reductions \( X \to U \) and \( X \to V \), there is a word \( W \) and reductions \( U \to W \) and \( V \to W \) such that the resulting diagram commutes. This might be a pushout in the rewrite category, but not necessarily.

Local confluence does not imply confluence, but the diamond property does. See [2] for a thorough treatment of these concepts.

Note, however, that our reductions have semantic content as well as syntactic. In our definitions of confluence, local confluence, and the diamond property, it is not enough that two derivations derive the same word; the resulting diagrams must also commute in the rewrite category. We say that two derivations \( X \overset{*}{\to} Y \) are equivalent if the composition of the 2-cells along the two paths are equal.

**Example 3.4.** One of the defining properties for monads, namely the left-hand diagram of (3.1), says that \( \mu \) as a reduction rule can be applied to the string \( T^3 \) in two ways to obtain \( T^2 \): one way as \( \mu T \) to the leftmost two occurrences of \( T \) (the left arrow of the diagram) and the other as \( T \mu \) to the rightmost (the top arrow), in both cases giving \( T^2 \). By applying \( \mu \) again to the two occurrences of \( T^2 \), we obtain a commutative diamond.

By (2.4), we can compose on the left and right with any strings in \( T^* \):

\[
\begin{array}{c c c}
T^{m+3} & \overset{T^{m+1} \mu T^n}{\longrightarrow} & T^{m+n+2} \\
\downarrow T^{m+1} & & \downarrow T^{m} \\
T^{m+n+2} & \overset{T^{m} \mu T^n}{\longrightarrow} & T^{m+n+1}
\end{array}
\] (3.2)

This says that any two reductions involving the rewrite rule \( \mu \) with overlapping redexes, applied anywhere in a string of length at least three, can be completed to a commutative diamond.

For nonoverlapping redexes, the diamond property always holds. This is a consequence of property (2.2) of 2-categories.

**Lemma 3.5.** Any two applications of rewrite rules with disjoint redexes can be applied in either order, and the resulting diagram commutes. That is, any well-typed diamond of the form

\[
\begin{array}{c c c}
P Q R S T & \overset{P Q R T}{\longrightarrow} & P Q R Y T \\
\downarrow P \sigma R S T & & \downarrow P \sigma R Y T \\
P X R S T & \overset{P X R T}{\longrightarrow} & P X R Y T
\end{array}
\] (3.3)

commutes, where \( P, Q, R, S, T, X, Y \) are 1-cells and \( \sigma : Q \to X \) and \( \tau : S \to Y \) are 2-cells.
Proof. It follows from property (2.2) of 2-categories that if \( \varphi : F \to G \) and \( \psi : M \to N \) are 2-cells and the following diagram is well-typed, then it commutes.

\[
\begin{array}{ccc}
F_M & \xrightarrow{F\psi} & FN \\
\downarrow \varphi M & & \downarrow \varphi N \\
G_M & \xrightarrow{G\psi} & GN
\end{array}
\]  

(3.4)

(In fact, (2.2) is equivalent to (2.3) and (3.4).) The diagram (3.3) is a special case of (3.4) with the following substitutions: \( PQ \) for \( F \), \( RST \) for \( M \), \( PX \) for \( G \), \( RYT \) for \( N \), \( P\sigma \) for \( \varphi \), and \( R\tau T \) for \( \psi \).

The property (2.1) of natural transformations is a stronger form of (3.4).

4. Applications

In this section we demonstrate the use rewrite categories in the verification of monad composition as presented by Eklund et al. \[7\]. We also give a characterization of monads and adjunctions in terms of rewrite categories.

The following lemma and its proof introduce our approach at a basic level.

**Lemma 4.1.** Let \( T^+ \) denote the set of nonnull strings of \( T \)'s and \( \mu \) the rewrite rule \( T^2 \to T \). The following are equivalent:

1. \( \mu \) satisfies the left-hand diagram of (3.1);
2. \( T \) is a terminal object in the rewrite category \( (T^+; \mu) \).

**Proof.** (1) \( \Rightarrow \) (2): Combining Lemma 3.5 with the observation (3.2), we have that the rewrite system consisting of the single rule \( \mu \) on strings \( T^n \) for \( n \geq 1 \) satisfies the diamond property and is therefore confluent. It follows that any diagram starting from \( T^n \), \( n \geq 1 \), and ending with the normal form \( T \) commutes. Moreover, there exists a reduction sequence from any such \( T^n \) to \( T \). Thus there is a unique morphism \( T^n \to T \) for \( n \geq 1 \), so \( T \) is a terminal object.

(2) \( \Rightarrow \) (1): Conversely, if \( T \) is a terminal object, then the left-hand diagram of (3.1) must commute, since there is a unique morphism \( T^3 \to T \). \( \square \)

4.1. Monads. Now we add the rewrite rule \( \eta \) to the mix. This rule can be used to introduce a new occurrence of \( T \) anywhere in the string. In the presence of \( \eta \), \( T \) is no longer a normal form, and in fact normal forms in the strict sense of string rewriting no longer exist. Nevertheless, \( T \) is still a terminal object. Moreover, \( T^0 = I \) can now be included, since there is an arrow \( I \to T \).

**Theorem 4.2.** Consider a rewrite category on 1-cells \( T^+ \) and rules \( \mu : T^2 \to T \) and \( \eta : I \to T \). The following are equivalent:

1. \( \mu \) and \( \eta \) satisfy (3.1);
2. \( T \) is a terminal object in the rewrite category \( (T^+; \mu, \eta) \).

**Proof.** (1) \( \Rightarrow \) (2): Suppose we have a rewrite system in which the rules can be classified as either bad rules or good rules (in our application, a rule is bad if it increases the length of the string, e.g. \( \eta \)). Call a derivation good if it uses only good rules. Suppose further that

\[\]
1. every pair of good reductions $X \rightarrow U$ and $X \rightarrow V$ can be completed to a diamond using only good reductions $U \rightarrow W$ and $V \rightarrow W$; and
2. every pair of reductions $X \rightarrow U$ and $X \rightarrow V$, good or bad, are confluent using only good derivations $U \ast \rightarrow W$ and $V \ast \rightarrow W$.

These conditions hold for $(T^*; \mu, \eta)$ under the assumption (1). We have already argued condition 1 for $\mu$ in Lemma 4.1. The right-hand diagram in (3.1) implies condition 2 by immediately inverting any application of $\eta$ whenever it is applied to a nonnull string. For example, consider applications of $\eta$ and $\mu$ to a substring $T^2$, where $\eta$ is applied between the two occurrences of $T$. The two redexes thus overlap.

The top arrow is good, but the left arrow is bad. However, using the right-hand diagram of (3.1), the diagram can be completed using only good arrows:

Now we argue that any system satisfying 1 and 2 is confluent. Let $X \ast \rightarrow U$ and $X \ast \rightarrow V$ be any two derivations involving good or bad rules. Starting from the apex $X$, move down the two derivations, adding good diamonds to the diagram in the case 1 and good confluent derivations in case 2. All new transitions added to the diagram are good. When done, there are no more exposed bad rules, and the diagram can be completed by filling in with good diamonds.

Thus any two derivations $T^n \ast \rightarrow T$ are confluent via good derivations, which must be of the form $T \rightarrow T$. But the only good derivation $T \rightarrow T$ is the identity. It follows that any two derivations $T^n \ast \rightarrow T$ are equivalent; in other words, $T$ is a terminal object.

$(2) \Rightarrow (1)$: Conversely, if $T$ is a terminal object of the rewrite category, then all diagrams starting with any $T^n$ and ending with $T$ must commute, in particular those of (3.1), the defining conditions for monads.

**Corollary 4.3.** Let $\sigma$ be a 2-functor from the free 2-category on one 0-cell, one 1-cell $T$, and two 2-cells $\mu, \eta$ to $\text{Cat}$, the 2-category of categories, functors, and natural transformations. Let $E$ be the equational theory on 2-cells induced by $\sigma$. The following are equivalent:

- The equations (3.1) are a logical consequence of $E$.
- $T$ is terminal in the rewrite category $(T^*; \mu, \eta)$ modulo $E$.
- The image $(\sigma(T), \sigma(\mu), \sigma(\eta))$ is a monad.

**4.2. Monad Composition.** Let $(P, \mu^P, \eta^P)$ and $(T, \mu^T, \eta^T)$ be monads on a category $C$ connected by a distributive law (or swapper in the terminology of [7]), which is a natural
transformation \( \theta : TP \to PT \) satisfying the following properties:

\[
\begin{array}{c}
\begin{array}{ccc}
TP^2 & \xrightarrow{\theta_P} & PT^2 \\
\downarrow{\mu^P} & & \downarrow{\mu^T} \\
TP & \xrightarrow{\theta} & PT \\
\downarrow{\mu^T} & & \downarrow{\mu^T} \\
T^2P & \xrightarrow{\eta^T} & PT^2
\end{array}
\end{array}
\]

\[
\begin{array}{c}
\begin{array}{ccc}
TP & \xleftarrow{\eta^T} & P \\
\downarrow{T\eta^P} & \xleftarrow{\theta} & \downarrow{PT} \\
T & \xleftarrow{\theta} & PT \\
\downarrow{\eta^P} & & \downarrow{PT^2}
\end{array}
\end{array}
\]

(4.1)

Distributive laws are discussed in depth in [4, §9.2]. A typical application is the construction of the complex algebra of an algebra, whose elements are sets of elements of the original algebra. Here \( P \) would be the powerset monad and \( T \) the term monad of some variety of algebras, and \( \theta \) takes a term of sets \( t(A_1, \ldots, A_n) \) and turns it into a set of terms \( \{t(a_1, \ldots, a_n) \mid a_i \in A_i, \ 1 \leq i \leq n\} \). These constructions are discussed in [5, 9]. Another example would be the combination of the additive and multiplicative monoid structures in semirings. Here \( P \) would be the finite powerset monad and \( T \) the free monoid construction.

In light of the theme of this note, it should be clear that the conditions in (4.1) are nothing more than a way to establish local confluence in the case of overlapping redexes between \( \theta \) and the monad operations. For example, the top rectangle of the left-hand diagram of (4.1) handles overlapping redexes involving \( \theta \) and \( \mu^P \).

The two monads \( P \) and \( T \) can be combined as follows. Define

\[
\mu^{PT} = \mu^P T \circ \mu^T \circ P \theta T : (PT)^2 \to PT \quad \eta^{PT} = \eta^P T \circ \eta^T : I \to PT.
\]

Then \( (PT, \mu^{PT}, \eta^{PT}) \) is again a monad [4, 7, 10]. We will verify this using Theorem 4.2. Most of the work is contained in the following lemma.

**Lemma 4.4.** Consider a rewrite category on 1-cells \( \{P, T\}^* \) in a 2-category with a single 0-cell and rules \( \mu^P : P^2 \to P, \eta^P : I \to P, \mu^T : T^2 \to T, \eta^T : I \to T \), and \( \theta : TP \to PT \), such that \((\mu^P, \eta^P)\) and \((\mu^T, \eta^T)\) both satisfy the monad axioms (3.1) (with appropriate substitutions). The following statements are equivalent:

1. \( \mu^P, \eta^P, \mu^T, \eta^T, \) and \( \theta \) satisfy the distributive laws (4.1).
2. \( PT \) is a terminal object in the rewrite category \((\{P, T\}^*; \mu^P, \mu^T, \eta^P, \eta^T, \theta)\).

*Proof.* (1) \( \Rightarrow \) (2): First we show that the rewrite system is confluent. Recall that a rule is *bad* if it increases length, *good* otherwise. The good rules are \( \mu^P \), \( \mu^T \), and \( \theta \), and the bad rules are \( \eta^P \) and \( \eta^T \).

Every pair of good reductions can be completed to a good confluent diagram. If the redexes do not overlap, this follows from Lemma 3.5. For redexes that overlap, all cases are covered by the left-hand diagram of (3.1) and the left-hand diagram of (4.1).

Given any derivation \( \pi : X \xrightarrow{*} Y \), possibly containing bad reductions, produce a new derivation \( \pi' \) as follows:

1. Extend the derivation to derive \( PT \).
2. Rearrange the resulting derivation \( X \xrightarrow{*} Y \xrightarrow{*} PT \) to obtain an equivalent derivation \( \pi' : X \xrightarrow{*} PT \) in which all the bad rules are applied after all the good rules.

Step 1 can be accomplished by first introducing an occurrence of \( P \) and/or \( T \) using \( \eta^P \) and \( \eta^T \) if necessary, then moving all occurrences of \( P \) to the left of all occurrences of \( T \) using \( \theta \), then collapsing the \( P \)'s using \( \mu^P \) and the \( T \)'s using \( \mu^T \).
Step 2 can be done without increasing the length of the derivation. For any bad reduction followed immediately by a good reduction, if the symbol introduced by the bad reduction is not part of the redex of the good reduction, then the two reductions can be switched by Lemma 3.5.

Otherwise, the symbol introduced by the bad reduction is part of the redex of the good reduction. There are only six ways this can happen:

\[
\begin{align*}
T & \xrightarrow{T_\eta T} T_2 \xrightarrow{\mu T} T \\
T & \xrightarrow{T_\eta T} T_2 \xrightarrow{\mu T} T \\
T & \xrightarrow{T_\eta P} T P \xrightarrow{\theta} PT
\end{align*}
\]

and symmetrically for \(P\). The first two are equivalent to the identity by the right-hand diagram of (3.1) and can be deleted. The last is equivalent to \(T_\eta T \xrightarrow{PT}\) by the right-hand diagram of (4.1). We can continue this process until there are no more bad reductions occurring before good reductions in the derivation.

If \(X\) contains at least one occurrence each of \(P\) and \(T\), then this must also be true of any string derived from \(X\), since all rules preserve this property. But then \(\pi'\) can contain no bad reductions at all. If it did, then the last reduction would be bad. But it is impossible to derive \(PT\) from such a reduction, since it would have to come from a string of length one, and no such string can be derived from \(X\).

By a similar argument, if \(X \in P^+\), then \(\pi'\) contains exactly one bad reduction to introduce \(T\), and it occurs last in the derivation. This last reduction must be of the form \(P \xrightarrow{P_\eta T} PT\). Similarly, if \(X \in T^+\), the last reduction of \(\pi'\) is of the form \(T_\eta P \xrightarrow{PT}\), and this is the only bad reduction in the derivation.

If \(X = I\), a similar argument shows that \(\pi'\) must be either

\[
I \xrightarrow{I_\eta} T \xrightarrow{T_\eta P} PT \\
I \xrightarrow{I_\eta} P \xrightarrow{P_\eta T} PT.
\]

(4.3)

Now suppose we are given derivations \(\pi : X \xrightarrow{\ast} U\) and \(\rho : X \xrightarrow{\ast} V\), possibly using both good and bad rules. To show confluence of \(\pi\) and \(\rho\), it suffices to show that \(\pi' : X \xrightarrow{\ast} PT\) and \(\rho' : X \xrightarrow{\ast} PT\) are confluent.

As argued above, if \(X\) contains at least one occurrence each of \(P\) and \(T\), then \(\pi'\) and \(\rho'\) are good. Thus we can complete them to a good commutative diagram by filling in with good commutative diamonds and (4.1). This process must terminate, since there is a fixed upper bound, quadratic in the length of \(X\), on the length of any good derivation from \(X\), since each good reduction strictly decreases the string in length or lexicographic order relative to \(P < T\) [2, Lemma 2.7.2].

If \(X \in P^+\), then as argued above, \(\pi'\) and \(\rho'\) are both of the form \(X \xrightarrow{\ast} P \xrightarrow{P_\eta T} PT\), where the prefixes \(X \xrightarrow{\ast} P\) contain no occurrence of \(T\). By Theorem 4.2, \(P\) is terminal in \((P^*; \mu P, \eta P)\), therefore the two prefixes \(X \xrightarrow{\ast} P\) are equivalent, and consequently so are \(\pi'\) and \(\rho'\). The argument is similar for \(X \in T^+\).

When \(X = I\), we need only observe that the two derivations (4.3) form a commutative diamond by Lemma 3.5.

Finally we show that \(PT\) is a terminal object. We have already argued that there is at least one derivation of \(PT\) from every \(X \in \{P,T\}^*\), so there is at least one arrow \(X \xrightarrow{\ast} PT\) in the rewrite category. To show that there is at most one, let \(\pi, \rho : X \xrightarrow{\ast} PT\) be any two derivations. By confluence, we can complete to a pair of equivalent derivations
\[ \pi', \rho' : X \rightarrow PT \rightarrow PT. \]

Rearranging the final portions \( PT \rightarrow PT \) of these two derivations by
Step 2 above, we obtain good derivations. But the only good derivation \( PT \rightarrow PT \) is the
identity, therefore \( \pi \) and \( \rho \) were already equivalent.

(2) \( \Rightarrow \) (1): If \( PT \) is a terminal object, then (4.1) holds, as all maximal paths in all
diagrams of (4.1) lead to \( PT \). \( \square \)

**Theorem 4.5** ([4, 7, 10]). Let \( (T, \mu^T, \eta^T) \) and \( (P, \mu^P, \eta^P) \) be monads connected by a
distributive law \( \theta : TP \rightarrow PT \) satisfying (4.1). Let \( \mu^{PT} \) and \( \eta^{PT} \) be defined by (4.2). Then
\( (PT, \mu^{PT}, \eta^{PT}) \) is a monad.

**Proof.** Consider the rewrite category\(^1\) \( \{P, T\}^*; \mu^P, \mu^T, \eta^P, \eta^T, \theta \). By Theorem 4.2, it suffices
to show that \( PT \) is a terminal object in the rewrite category \( (PT)^*; \mu^{PT}, \eta^{PT} \). By Lemma
4.4, we know that it is terminal in the category \( (P, T)^*; \mu^P, \mu^T, \eta^P, \eta^T, \theta \), thus for any
\( n \geq 0 \), there is exactly one arrow \( (PT)^n \rightarrow PT \) in that category. It follows that there is at
most one arrow \( (PT)^n \rightarrow PT \) in the subcategory \( (PT)^*; \mu^{PT}, \eta^{PT} \). But there is also at
least one, since we can derive \( PT \) from \( (PT)^0 = I \) by the rule \( \eta^{PT} \) and from \( (PT)^n \) for \( n \geq 1 \)
by \( n - 1 \) applications of the rule \( \mu^{PT} \). \( \square \)

4.3. Adjunctions. As a final application, we give a characterization of adjoint functors in
terms of rewrite categories.

Recall that a functor \( F : \mathcal{C} \rightarrow \mathcal{D} \) is a left adjoint of another functor \( G : \mathcal{D} \rightarrow \mathcal{C} \) if there are
natural transformations \( \eta : I \rightarrow GF \) and \( \varepsilon : FG \rightarrow I \), the unit and counit of the adjunction,
respectively, satisfying

\[
\begin{array}{ccc}
GFG & \xrightarrow{G\varepsilon} & G \\
\downarrow \eta G & \downarrow \mathrm{id} & \downarrow \mathrm{id} \\
\end{array}
\]

\[
\begin{array}{ccc}
F & \xrightarrow{F\eta} & FGF \\
\downarrow \varepsilon F & \downarrow \mathrm{id} & \downarrow \mathrm{id} \\
\end{array}
\] \tag{4.4}

**Theorem 4.6.** Consider a 2-category on two 0-cells \( \mathcal{C}, \mathcal{D} \) generated by 1-cells \( F : \mathcal{C} \rightarrow \mathcal{D} \) and
\( G : \mathcal{D} \rightarrow \mathcal{C} \) and 2-cells \( \eta : I \rightarrow GF \) and \( \varepsilon : FG \rightarrow I \). The following statements are equivalent:

1. \( \eta \) and \( \varepsilon \) satisfy (4.4).
2. \( F \) is a terminal object in the rewrite category \( (F(GF))^*; \eta, \varepsilon \) and \( G \) is a terminal object
   in the rewrite category \( (G(FG))^*; \eta, \varepsilon \).

**Proof.** (1) \( \Rightarrow \) (2): By arguments similar to the proof of Theorem 4.2 and Lemma 4.4, in any
derivation \( \pi : X \rightarrow Y \), all applications of \( \eta \) (the bad rule) can be moved after all applications of
\( \varepsilon \) (the good rule), by either Lemma 3.5 in the case of disjoint redexes or (4.4) in the case of
overlapping redexes. The resulting equivalent derivation \( \pi' : X \rightarrow Y \) is no longer than \( \pi \).

Now if \( \pi' \) is of the form \( F(GF)^n \rightarrow F \), there can be no application of \( \eta \), because the last
\( \eta \) would produce a string of length at least 2. But then all redexes are redexes of \( \varepsilon \), therefore
are disjoint and can be done in any order by Lemma 3.5. Thus all derivations of the form
\( F(GF)^n \rightarrow F \) are equivalent. Moreover, there exists a derivation \( F(GF)^n \rightarrow F \) consisting of
\( n - 1 \) applications of \( \varepsilon \). Thus \( F \) is terminal in \( (F(GF))^*; \eta, \varepsilon \). A symmetric argument shows that
\( G \) is terminal in \( (G(FG))^*; \eta, \varepsilon \).

\(^1\)At the risk of ambiguity, we are overloading the symbols \( P, \mu^P \), etc. to refer to the components of the
rewrite category as well as their images in \( \mathcal{Cat} \).
(2) ⇒ (1): In (4.4), there is a unique morphism to $G$ in the left-hand diagram and a unique morphism to $F$ in the right-hand diagram, thus the diagrams commute.

**Corollary 4.7.** Let $\sigma$ be a 2-functor from the free 2-category on two 0-cells $C, D$, two 1-cells $F : C \to D$, $G : D \to C$, and two 2-cells $\eta : I \to GF$, $\varepsilon : FG \to I$ to $\text{Cat}$. Let $E$ be the equational theory on 2-cells induced by $\sigma^{-1}$. The following are equivalent:

- The equations (4.4) are a logical consequence of $E$.
- $F$ is terminal in the rewrite category $(F(GF)^*; \eta, \varepsilon)$ modulo $E$ and $G$ is terminal in the rewrite category $(G(FG)^*; \eta, \varepsilon)$ modulo $E$.
- $\sigma(F)$ and $\sigma(G)$ are adjoint functors with $\sigma(F)$ the left adjoint and $\sigma(G)$ the right adjoint with unit $\sigma(\eta)$ and counit $\sigma(\varepsilon)$.

5. Conclusion

It is clear that these techniques are related to the notion of *free adjunction* and *free monad* as presented by Schanuel and Street [15]. A free adjunction is a certain 2-category that is initial among all adjunctions and characterizes their equational theory. For the future, we would like to explore these connections further and perhaps develop a notion of rewriting for 2-categories and higher.

**Acknowledgments**

Thanks to Riccardo Pucella for valuable comments.

**References**


