Abstract

Explicit substitutions are an important tool for the efficient implementation of the normalization of terms in programs that manipulate data with binders, such as theorem provers or type checkers. We explore here the design space between the λσ-calculus [Abadi et al. 1991] and the suspension-calculus [Nadathur and Wilson 1998] by showing how to go from one to the other, in various small steps. This gives an intuition about the performance advantages of the suspension-calculus as well as provides various alternatives.

1. Introduction

The substitution operation at the core of most formal languages, sometimes written using a notation such as \( e_1[e_2/x] \), is often taken for granted but can be surprisingly tricky to formalize correctly. This motivated the development of explicit substitutions calculi such as λσ [Abadi et al. 1991] where substitution is not taken as a black box, but is instead decomposed into several primitive steps, reified explicitly as part of the syntax of terms, along with rewrite rules.

Explicit substitutions calculi are especially useful for implementing theorem provers or type checkers. In both applications, one of the driving motivations for their use is their ability to combine substitutions of various variables into a single term traversal. This makes a significant difference to the performance of operations such as term normalization and higher order unification [Liang et al. 2004].

Among the various substitutions calculi, the ones that are able to combine individual substitutions into a single term traversal fall into two large families: the variants of Abadi’s λσ and the various flavors of the suspension-calculus [Nadathur and Wilson 1998]. The first is a theoretical formalism that aims to bring out the core concepts as cleanly as possible, whereas the second was developed by one the de Bruijn index of every variable reference; and σ₁ □ σ₂ is the composition of the application of the substitution σ₁, followed by the application of the substitution σ₂. The cons operator □ binds tighter than the composition operator □. The intuition underlying these substitution terms is well explained in the original article [Abadi et al. 1991].

Notice that this syntax represents non-zero de Bruijn indices via \( \#0 \) and is often taken tightly than the \( \#0 \) operator which replaces \( \#0 \) with \( t \). In subscripts, \( \#0 \) is a de Bruijn index referencing the most recently introduced variable; \( t_1 t_2 \) is a function application; and \( t[\sigma] \) is a closure or suspension which stands for the substitution \( \sigma \) applied to the term \( t \).

In substitutions, id is the identity substitution which does not substitute anything; \( t \cdot \sigma \) is called a cons and is a substitution which replaces \( \#0 \) with \( t \) and applies \( \sigma \) to the other variables; \( t \cdot \sigma \) is a function application; \( \lambda t \) is a lambda abstraction; and \( t[\sigma] \) is a closure or suspension which stands for the substitution \( \sigma \) applied to the term \( t \).

In substitutions, id is the identity substitution which does not substitute anything; \( t \cdot \sigma \) is called a cons and is a substitution which replaces \( \#0 \) with \( t \) and applies \( \sigma \) to the other variables; \( ↑ \) is pronounced shift and is the substitution which simply increments by one the de Bruijn index of every variable reference; and \( \sigma_1 \cdot \sigma_2 \) is the composition of the application of the substitution \( \sigma_1 \), followed by the application of the substitution \( \sigma_2 \). The cons operator □ binds tighter than the composition operator □. The intuition underlying these substitution terms is well explained in the original article [Abadi et al. 1991].

The syntax is split into 3 parts: the \( \beta \) rule which introduces suspensions, the reading rules which describe how to propagate substitutions down the terms, and the merge rules which describe how to reduce compositions to simpler substitution expressions. It can be shown that all the rules together correspond to the classical \( \beta \)-reduction rule.

In [Curien et al. 1996], the authors investigate the confluence properties of this calculus.

3. Implementing Explicit Substitutions

That formalization of \( \lambda \sigma \)-calculus does not lend itself immediately to an implementation, for the following reasons:

\[
\begin{align*}
  (\beta) & \quad (\lambda t_1) t_2 \rightsquigarrow t_1[t_2 \cdot \text{id}] \\
  (\beta_0) & \quad \#0[t \cdot \sigma] \rightsquigarrow t \\
  (\beta_1) & \quad \#0[\text{id}] \rightsquigarrow \#0 \\
  (\beta_2) & \quad (t_1 t_2)[\sigma] \rightsquigarrow t_1[\sigma] t_2[\sigma] \\
  (\beta_3) & \quad (\lambda t)[\sigma] \rightsquigarrow \lambda t[\#0 \cdot (\sigma \cdot \uparrow)] \\
  (\beta_4) & \quad t[\sigma_1][\sigma_2] \rightsquigarrow t[\sigma_1 \cdot \sigma_2] \\
  (\text{id}) & \quad \text{id} \cdot \sigma \rightsquigarrow \sigma \\
  (\text{id}) & \quad \text{id} \cdot \text{id} \rightsquigarrow \uparrow \\
  (\text{id}) & \quad \uparrow \circ (t \cdot \sigma) \rightsquigarrow \sigma \\
  (\text{cons}) & \quad (\sigma_1 \cdot \sigma_2) \cdot \sigma_3 \rightsquigarrow \sigma_1 \cdot (\sigma_2 \cdot \sigma_3) \\
  (\text{bind}) & \quad (t \cdot \sigma_1) \cdot \sigma_2 \rightsquigarrow t[\sigma_2] \cdot (\sigma_1 \cdot \sigma_2)
\end{align*}
\]
In Fig. 2, we see that compose is only explicitly needed to handle terms of the form \( t[\sigma_1][\sigma_2] \). Looking further at the rules, we can see that such terms can be introduced only in the following cases:

1. One of the reading rules when a subterm is already itself a suspension.
2. The \( \beta \) rule when \( t_1 \) is itself a suspension.

When performing the typical normalization by reduction of outermost redexes, the reading rules should only be applied to terms that haven’t been visited yet, so we have the property that there should never be a suspension \( t'[\sigma'] \) inside the \( t \) of another suspension \( t[\sigma] \). In other words, the first case should never happen. Similarly, in the normalization case, the second case should only occur if the suspension \( t_1 \) was pushed earlier from outside the \( \lambda \) by the \( \text{rwm} \) rule.

As it turns out, we can add a special \( \beta' \) rule to handle the second case without resorting to compose:

\[
(\beta') \; (\lambda t_1)[\sigma_1] t_2 \rightsquigarrow t_1[\text{cons} t_2 \ id]
\]

This rule was already suggested in [Abadi et al. 1991] and a similar rule can be found in the \( \lambda^* \) calculus [Accattoli and Kesner 2010]. Note that the \( \beta' \) rule does not rule out the need of a compose function if one of the other 4 functions makes use of it internally.

3.2 Implementation Suggested for \( \lambda\sigma \)-calculus

In their paper, [Abadi et al. 1991] propose an implementation strategy for weak head normalization. We can define our 5 functions to correspond to their implementation strategy as shown in Fig. 3.

4. \( \lambda\sigma_0 \): Restricted \( \lambda\sigma \)-calculus

The \( \lambda\sigma \)-calculus presented so far has a very efficient implementation of compose but at the cost of a rather slow and complex implementation of lookup. Furthermore a large part of the computation performed in lookup risks being executed repeatedly for the same substitution applied to several variable occurrences.

Another problem is that sequences of \( \uparrow \) tend to arise in many scenarios. For example, it can be shown [Abadi et al. 1991] that the normal form of a substitution in the \( \lambda\sigma \)-calculus is of the form \( t_1 \ldots t_n \cdot (\uparrow \circ \ldots \circ \uparrow) \), Thus most substitutions will contain a sequence of \( \uparrow \), and we would like to represent them more efficiently.

For these reasons, in Twelf [Pfenning and Schürmann 1999], the implementors decided to use a restricted form of the \( \lambda\sigma \)-calculus which we call \( \lambda\sigma_0 \) where the composition is not part of the syntax of substitutions but is implemented as a function instead and where \( \uparrow \) is extended to \( \uparrow^\circ \), so multiple shifts can be combined into a single one.
The corresponding syntax and instantiation of the parameters of the reading rules are shown in Fig. 4. As can be seen, the syntax is pleasantly simplified and the implementation is also streamlined. Substitutions are now represented as singly-linked lists where the terminating “nil” element carries the number of shifts to apply to the other variables.

lookups are now efficient, but at the cost of more expensive compose. Furthermore, compose is now used internally for lift, so the rule $\beta$ is not sufficient to make the performance of compose irrelevant.

### 5. $\lambda\sigma_1$: Shifting Lazily

The performance of the $\lambda\sigma_0$-calculus tends to suffer from the following two issues:

- The call to compose in lift will always fall through to the last case of the compose function, which recurses until the end of the substitution. So it takes time proportional to the length of the substitution, and since lift increases the length of the substitution by 1, a sequence of $N$ applications of lift has an $O(N^2)$ complexity. It would be preferable to try and push those $\uparrow$ to the leaves more lazily, so we can combine them along the way.

- Another related problem comes when we need to compute something of the form $(t_1 \cdot \sigma_1 \uparrow_{n_1}) \circ t_2 \cdot \sigma_2$. Twelf will reduce it as follows:

\[
(t_1 \cdot \sigma_1 \uparrow_{n_1}) \circ t_2 \cdot \sigma_2 \\
\leadsto (t_1 \uparrow_{n_1} \circ (\sigma_1 \uparrow_{n_1})) \circ t_2 \cdot \sigma_2 \\
\leadsto (t_1 \uparrow_{n_1} \circ \cdots) \circ t_2 \cdot \sigma_2 \\
\leadsto t_1 \uparrow_{n_1} (t_2 \cdot \sigma_2) \circ (\cdots \circ t_2 \cdot \sigma_2) \\
\leadsto t_1 \uparrow_{n_1} (t_2 \cdot \sigma_2) \cdot (\cdots \circ t_2 \cdot \sigma_2) \\
\leadsto t_1 \uparrow_{n_1} (t_2 \cdot \sigma_2) \\
\leadsto t_1 \uparrow_{n_1} (t_2 \cdot \sigma_2) \\
\leadsto t_1 \cdot \sigma_1 \uparrow_{n_1} \circ t_2 \cdot \sigma_2 \\
\leadsto t_1 \cdot \sigma_1 \uparrow_{n_1} \circ t_2 \cdot \sigma_2
\]

whereas we would want to get rid of $t_2$ right from the start by reducing in a different order, such as:

\[
(t_1 \cdot \sigma_1 \cdot \uparrow_{n_1}) \circ t_2 \cdot \sigma_2 \\
\leadsto (t_1 \cdot \sigma_1) \cdot (\uparrow_{n_1} \circ t_2) \circ \sigma_2 \\
\leadsto (t_1 \cdot \sigma_1) \cdot (\cdots \circ t_2 \cdot \sigma_2) \\
\leadsto t_1 \cdot \sigma_1 \cdot (t_2 \cdot \sigma_2) \cdot (\cdots \circ t_2 \cdot \sigma_2)
\]

We can solve those two problems by using the following tweak to Twelf’s approach: replace the $\uparrow$ which just does a shift, with $\uparrow$ which is a new syntax whose meaning is equivalent to $\sigma \circ \uparrow$ in the $\lambda\sigma$-calculus. Fig. 5 shows the resulting syntax and implementation.

Notice that we had to re-introduce an id substitution. More importantly, notice that the rule for compose $\sigma_1 (\uparrow_{n_1} \sigma_2)$ and the rule for compose $(t \cdot \sigma_1) \sigma_2$ overlap, and we use the definition’s ordering to make sure that we always use the first, which hoists the shifts outward, in preference to the second, which pushes the shifts deeper and duplicates them.

One way to think about it is that the $\lambda\sigma_1$-calculus tries to keep the shifts close to the top of the substitution, so they can quickly cancel out a cons. This is done in the second case of the compose function. In contrast Twelf’s approach always pushes the shifts all the way to the bottom of the substitutions, so we have to do more work before we can cancel a cons with a shift. Of course, there is no free lunch: we still have to propagate the shifts to the very bottom sooner or later, but we do it in the lookup rules instead, where the parameter $o$ keeps track of how many shifts we have encountered along the way.

### 6. $\lambda\sigma_2$: Fast Lookup of Free Variables

One inefficiency in $\lambda\sigma_1$ is the treatment of lookup for free variables that are outside the scope of the substitution. More specifically, every substitution has a length, which is the number of cons elements. Looking at the lookup’ function in Fig. 5, one can see that any variable reference $\#i$ whose $i$ is larger than this length will end up in the id case and be turned into $\#(i-l+o)$ where $l$ is the length of the environment and $o$ is the number of shifts in the substitution.

In other words, for any $\#i$ where $i$ is bigger than the length of $\sigma$, lookup’ will traverse the complete $\sigma$, only to collect the number of cons and the number of shifts.

For example, the term $(\lambda \#3) t_1$ will be reduced to $\#3[t_1 \cdot id]$. The result is $\#2$ which corresponds to collecting 1 cons and 0 shift. As another example the term $(\lambda \#3) t_1$ will be reduced to $\lambda \#3[\#0 \uparrow_{1} t_1 \cdot id]$. The result is $\lambda \#2$ which corresponds to collecting 2 cons and 1 shift.

The $\lambda\sigma_2$-calculus addresses this inefficiency by keeping track of the length and the total number of shifts in substitutions. More specifically, substitutions are now defined as triplets $(o, n, e)$ where $o$ is the length of $e$, $n$ is the number of shifts in it, and $e$ is the “raw” substitution, which has the same shape as the substitutions of $\lambda\sigma_1$. The syntax and implementation of the $\lambda\sigma_2$-calculus can be found in Fig. 6. The lookup’ function is identical to the one for $\lambda\sigma_1$, except that $\uparrow id$ is replace by $(0, 0, \uparrow id)$.

The way $\lambda\sigma_2$-calculus keeps track of the length of the substitution and the number of shifts might not be obvious to readers unfamiliar with the suspension calculus. The id substitution obviously has 0 cons and 0 shifts. The cons function increases by one the variable $o$ because one more term is added to the substitution. Recall that this rule is triggered by the $\beta$-rule. The function lift, in-
\[ \sigma := (\text{id}, n, l, e) \]  
\[ e := t \cdot e \uparrow_n e \]

\[ \text{id} = (0, 0, \text{id}) \]
\[ \text{cons } t (\text{id}, n, l, e) = (\text{id} + 1, n, l, t \cdot e) \]
\[ \text{lift } (\text{id}, n, l, e) = (\text{id} + 1, n, l + 1, \#(\text{id} \uparrow_n e) \]

**lookup #\(i\) (\text{id}, n, l, e) =**  
if \((i \geq \text{id})\) then  
\((#(i - \text{id} + n))\)  
else  
\((-\text{id}) \#(n)\) 
\[ \text{compose } \sigma \text{id} = \sigma \]
\[ \text{compose } (\text{id}, n, l, e) = (\text{id}, n, l', \text{id}(\text{compose } e_1 e_2)) \]
\[ \text{where } \text{ol'} = \text{id} + \text{max}(0, (\text{id} - \text{n})) \]
\[ \text{nl'} = \text{id} + \text{max}(0, (\text{nl} - \text{id})) \]
\[ \text{compose } (\text{id}, n, l, \uparrow \text{e}) = (\text{id}, n, l - 1, \text{e} - 1, \text{nl} - 1, \text{id}) \]
\[ \text{compose } (\text{id}, n, l, \text{n} \cdot \text{e}) = (\text{id}, n, l - 1, \text{nl} - 1, \text{n} \cdot \text{nl} - 1, \text{id}) \]
\[ \text{compose } (\text{id}, n, l, \text{n} \cdot \text{e}) = (\text{id}, n, l - 1, \text{nl} - 1, \text{nl} - 1, \text{id}) \]
\[ \text{where } \text{ol'} = \text{id} + \text{max}(0, (\text{id} - \text{n})) \]
\[ \text{nl'} = \text{id} + \text{max}(0, (\text{nl} - \text{id})) \]

\[ * \text{Figure 6. Syntax and functions of the } \lambda \sigma_2 \text{-calculus} * \]

creases both \(\text{id}\) and \(\text{n}l\) by one because it adds the term \(\#0\) at the front of the substitution and lifts the remaining part of the substitution. This rule is triggered by the \(r_i\)-rule when substitution go under a lambda.

The tricky part is how compose computes the new \(\text{id}\) and \(\text{n}l\) for the substitution it generates. When a \text{shift} cancels a \text{cons}, compose calls itself recursively with substitutions that contains one fewer \text{shift} and one fewer \text{cons}. But when compose creates a new substitution, it computes its new length to \(\text{id}\) and \(\text{n}l\) and its new number of \text{shifts} to \(\text{id}\) and \(\text{n}l\) are used to adjust the indices of the free variables. Also, compose is used for terms like \((\text{id}, n, l, e) \cdot \text{id}(\text{id})\).

\[ i > \text{id} \]

After the first substitution is applied, the variable now has the index:
\[ i' = i - \text{id} + \text{n}l \]

Now when the second substitution is applied, in order to be free the index \(i'\) must again be greater than the length of \(\text{nl}\). So we have
\[ i' = i - \text{id} + \text{n}l + 1 > \text{nl} \]

Because the composition of \(\text{id}\) and \(\text{n}l\) must give the same results as applying both substitution in order, its \(\text{id}\) is \(\text{max}(\text{id}, \text{nl} + \text{id} - \text{nl} - 1)\) which can be written \(\text{id} + \text{max}(0, \text{id} - \text{nl})\). The \text{max} function is used because both equations 1 and 3 must be satisfied.

For the parameter \(\text{nl}'\), we must have that the new variable index \(\text{id}'\) after the second substitution \(\sigma_2\) satisfies
\[ \text{id}' = i - \text{id} + \text{n}l + 1 \text{nl} + \text{nl} = i - \text{id} + \text{n}l + 1 \]

Again, the equation is justified because the composition of \(\text{id}\) and \(\text{n}l\) must give the same results as applying both substitution in order.

By having \(\text{id}' = \text{id} + \text{max}(0, \text{id} + \text{nl} - \text{id} - 1)\), it is easy to check that \(\text{n}l' = \text{n}l + \text{max}(0, \text{n}l - \text{id})\).

The strength of \(\lambda \sigma_2\)-calculus comes from the fact that we can compute \(\text{n}l\) and \(\text{id}\) when traversing a term even in the presence of substitutions composition and therefore always lookup the free variables of a term in constant time. We now turn to the suspension-calculus to improve the lookup of bound variables.

### 7. The Suspension Calculus

The reader familiar with the suspension calculus will notice a strong resemblance to the \(\lambda \sigma_2\)-calculus. To get to the suspension calculus we need to perform one last optimization. We would like to be able to lookup bound variable efficiently, without having to traverse the substitution. In \(\lambda \sigma_2\)-calculus we need to do so for two reasons. First, for a bound variable \(\#i\), it is impossible to locate directly the position of the \(i\)-th \text{cons} element in the data structure of \(\sigma\). Indeed, \(t \cdot \sigma\)-terms are mixed with \(\sigma \cdot \tau\)-terms. Second, even if we had a way to directly access the \(i\)-th \text{cons} element, we need to collect the number of \text{shifts}. The only way to do so is by traversing the list.

To lift the first restriction, the terms \(\tau\) can be made more compact by having each step be a combination of a \text{cons} with some number of \text{shifts}. More specifically, let’s consider the typical shape of an environment \(\tau\) in the \(\lambda \sigma_2\)-calculus:
\[ \text{\tau}_{n_1} (t_1 \cdot \text{\tau}_{n_2} (t_2 \cdot \cdots \cdot \text{\tau}_{n_m} \text{id})) \]

We can replace every \(\text{\tau}_{n} (t \cdot \sigma)\) with a new syntax \((t, n) :: \sigma\). This syntax merges \(t \cdot \sigma\) and \(\sigma \cdot \tau\) together. Also, considering the way \text{lookup} is implemented, we can see that the final \(\text{\tau}_{n_m} \text{id}\) is not really needed, since either \text{lookup} only uses \(\text{id}\) or \text{lookup}'\(\text{id}\) necessarily stops before reaching the end. We can replace \(\text{\tau}_{n_m} \text{id}\) by a constant \(\text{nil}\).

\[ (t_1, \text{\tau}_{n_1}) :: (t_2, \text{\tau}_{n_2}) :: \cdots :: \text{\tau}_{n_m} \text{id} \]

We now have a substitution represented as a list. Now \text{lookup} can be implemented efficiently. Not only does it skip the \(O(N)\) traversal for free-variables, but the traversal can use a simple indexation operation to directly fetch the \(n^\text{th}\) element. \(\tau\) could be implemented with any standard data structure such as an array or a balanced tree.

To lift the second restriction, the suspension calculus makes one more tweak to that representation: in order to avoid having to count all the \text{shifts} by traversing \(\tau\), the \((t, n) :: \sigma\) representation stores in \(n\) the difference between the number of shifts and the variable \(\text{id}\) of the substitution. In other words, we have the following equivalence:
\[ (t_1, \text{\tau}_{n_1}) :: (t_2, \text{\tau}_{n_2}) :: \cdots :: (t_m, \text{\tau}_{n_m}) :: \text{\tau}_{n_1} \text{id} \]

So to get the total number of \text{shifts} for the variable \#i, we must compute \(\text{\tau}_{n_1} \text{id} \cdot \tau_{n_1 - n_2} (t_2 \cdot \cdots \cdot \tau_{n_m - n_{m-1}} \text{id}) \cdot \text{id}) \]

The main difference is the absence of the \(\text{\tau}_{n}\) form. This form is used in the suspension-calculus as a shorthand for \((\#0, n +\)
\[
\sigma ::= \text{nil}, (\text{nl}, e), (\text{nl}, (\text{nl} + 1), e) \\
\text{id} ::= (0, 0, \text{nil}) \\
\text{cons } t (\text{nl}, e) \to (\text{nl} + 1, \text{nl}, t, \text{cons } (\text{nl}, e)) \\
\text{lift } (\text{nl}, e) \to (\text{nl} + 1, \text{nl} + 1, (\# t, \text{nl} + 1) : e) \\
\text{lookup} # i (\text{nl}, e) = \\
\begin{cases}
\text{if } (i \geq \text{nl}) & \text{then} \\
\# (i - \text{nl} + 1) & \text{else}
\end{cases}
\]

\[
\begin{align*}
\text{compose } & (\text{nl}, \text{nl}_1, e_1) (\text{nl}_2, \text{nl}_2, e_2) = \\
& (\text{nl}', \text{nl}'', \text{comp}_p, e_1 \text{nl}_1 \text{nl}_2 e_2) \\
\text{where} & \text{nl}' = \text{nl}_1 * \{0, (\text{nl}_2 - \text{nl}_1)\} \\
& \text{nl}'' = \text{nl}_2 * \{0, (\text{nl}_1 - \text{nl}_2)\} \\
\text{comp}_p & e_1 \text{nl}_1 0 \text{nl} = e_1 \\
\text{comp}_p & 0 \text{nl} \text{nl}_2 e_2 = e_2 \\
\text{comp}_p & \text{nil} \text{nl} \text{nl}_2 (t, n) : e_2 = \\
& \text{comp}_p \text{nil} (\text{nl}_1 - 1) (\text{nl}_2 - 1) e_2, \\
& \text{if } \text{nl}_1 \geq 1 \\
\text{comp}_p & (t_1, n_1) : e_1) \text{nl}_1 \text{nl}_2 (t_2, n_2) : e_2 = \\
& \text{comp}_p (t_1, n_1) : e_1) (\text{nl}_1 - 1) (\text{nl}_2 - 1) e_2 \\
& \text{if } \text{nl}_1 > n_1 \\
\text{comp}_p & (t_1, n_1) : e_1) \text{nl}_1 \text{nl}_2 (t_2, n_2) : e_2 = \\
& \text{let } e' = \text{comp}_p e_1 n_1 \text{nl}_2 (t_2, n_2) : e_2 \\
& \text{in } (t_1 (\text{nl}_2, \text{nl}_2, (t_2, n_2) : e_2), n_2 + \text{max}(0, (n_1 - \text{nl}_2))) : e' \\
\end{align*}
\]

1), which is what we use here in the \text{lift} function instead. The advantage of using a special form like \text{\text{\&\&}n} is to distinguish the case where a variable is replaced with another variable from the case where a variable’s index is simply renumbered. This difference is significant when we care to preserve information such as the location of the variable in the source code.

The reader will probably have noticed the absence of a \text{compose} function, which simply reflects the fact that this calculus relies on the \text{\text{\&\&}t} rule to avoid most uses of \text{compose}. And it comes with its own additional ad-hoc rule, hidden within the \text{lookup} function: before returning a new suspension shifted by \text{nl} = \text{n}, the function checks to see if the base form \text{t} is itself a suspension, in which case it performs the composition of the base substitution with the shift by \text{nl} - \text{n}, to avoid the need for a complete implementation of \text{compose}.

### 7.2 Simplified Suspension Calculus

Of course, if needed, we can define \text{compose}, for example by taking the rewrite rules presented in [Gacek and Nadathur 2007] and implementing them as a function as shown in Fig 8.

We can see in Fig 8 that compose is similar to the one defined for \text{\lambda\sigma}. The resulting composition has the same \text{nl}' and \text{nl}". The first two cases of \text{compose} correspond to the \text{id} case. The two other cases when a \text{shift} cancels a \text{cons}. Remember that nil replaces \text{n} \text{id} so the substitution can still represent \text{shifts} in the presence of \text{nil}. The last one corresponds to the last case of compose in \text{\lambda\sigma}, where the second substitution is distributed over the cons of the first.

### 8. Related Work

This is not the first time explicit substitutions calculi are compared. For example, Gacek and Nadathur [Gacek and Nadathur 2007] provide a formal mapping between the simplified suspension calculus and \text{\lambda\sigma} as well as many other calculi. They also provide a full treatment of the properties of the simplified suspension calculus and their mathematical proofs. Our intention here is to try and give a more intuitive and pragmatically motivated mapping, going into several motivated steps instead of giving a formal mapping.

### 9. Conclusion

In this paper we showed how the suspension calculus can be viewed as an optimized implementation of the original \text{\lambda\sigma}. We hope this can motivate the use of the suspension calculus as an implementation tool even though it is not easy to figure out how it works based on its own original rules. Indeed, like most optimized algorithms, trying to understand the inner working of the suspension calculus directly can be hard. It is simpler to understand \text{\lambda\sigma} and how to proceed to optimize it.

### 10. Acknowledgment

We would like to thank the reviewers for giving us various pointers to similar work in the literature.

### References


