The many disguises of accumulation

Eerke Boiten

University of Nijmegen, Dept. of Informatics

Abstract

Several descriptions of basically one transformation technique, viz. accumulation, are compared. Their basis, viz. the associativity and the existence of a neutral element inherent in a monoid, is identified.

Keywords

transformational programming, factorial, fast reverse, accumulation, continuations, lambda abstraction, generalisation, tail recursion, implementation of lists.

This research has been sponsored by the Netherlands Organisation for Scientific Research (NWO), under grant NF 63/62-518 (the STOP — Specification and Transformation Of Programs — project).
1 Introduction

One of the first program transformations that appeared in the literature was the accumulation transformation. The transformation is now classic, although not everyone may know it under exactly this name.

In this note, I try to relate several descriptions of this program transformation technique. In a purely algebraic view, it is the exploitation of the properties of a monoid. In literature, it can be found under the headings of recursion removal [Str71, WS73, AS78] or - simplification [Par90], generalisation of output terms [Aub75], data structures representing continuations [Wan80], rebracketing [BW82], accumulation [Bir84a], lambda abstraction or higher order generalisation [Pet87, PS87], or even a novel implementation of lists [Hug86]. It falls under the generalisation tactic in Feather’s survey paper [Fea87], but is not mentioned explicitly there.

I will start by showing the two “famous examples” of accumulation. Then I show the basis of the technique with the monoid. Finally, I comment on various papers by showing how they describe accumulation and what other transformations are possible using the described technique.

The examples come from many sources, each with their own variant of a functional notation, which is in some cases the clearest notation to express their ideas in. Therefore, I ask to be excused for the variations in notation throughout this note.

Both juxtaposition and bracketing are used for function application; \( \circ \) denotes function composition; \( \oplus \) always denotes a binary infix operator, with unit element \( 1_\oplus \).

2 Famous example \# 1: reverting lists

Assuming we are dealing with cons-lists over a (here, irrelevant) type \( \alpha \), defined by:

\[
\begin{align*}
[] : \text{List}(\alpha) \\
x : \alpha, l : \text{List}(\alpha) & \quad (x + l) : \text{List}(\alpha)
\end{align*}
\]

and, additionally, a function \( \oplus \) (append) defined by:

\[
\begin{align*}
[] \oplus l & = l \\
(a + l_1) \oplus l_2 & = a + (l_1 \oplus l_2)
\end{align*}
\]

i.e., the use of \( l_1 \oplus l_2 \) forces iteration over \( l_1 \). Thus, the reverse of a list can be defined
by:

\[
\begin{align*}
\text{rev}[\ ] &= [ ] \\
\text{rev}(a + l) &= (\text{rev } l) \mathbin{\underline{\mathbin{\underline{\mathbin{+}}}}}(a + [ ])
\end{align*}
\]

which is (given this definition of \(\mathbin{\underline{\mathbin{\underline{+}}}}\)) an algorithm quadratic in the length of the argument. By some, as of yet unknown, motive we are led to define

\[
h l_1 l_2 = (\text{rev } l_1) \mathbin{\underline{\mathbin{\underline{+}}}} l_2.
\]

Obviously, then

\[
\text{rev } l = h l[ ]
\]

and also

\[
\begin{align*}
h[ ]l_2 &= l_2 \\
h(a + l_1)l_2 &= h l_1(a + l_2)
\end{align*}
\]

which constitutes a linear algorithm for \text{rev}.

### 3 Famous example \# 2: accumulator version of factorial

We are given the usual definition of the factorial function \text{fact}, i.e.

\[
\begin{align*}
\text{fact}(0) &= 1 \\
\text{fact}(n) &= n \times \text{fact}(n - 1) \quad \text{for } n \geq 1
\end{align*}
\]

By generalising \(\text{fact}(n)\) to \(k \times \text{fact}(n)\) we obtain:

\[
\text{fact}'(k, n) = k \times \text{fact}(n)
\]

and thus,

\[
\text{fact}(n) = \text{fact}'(1, n)
\]

and also

\[
\begin{align*}
\text{fact}'(k, 0) &= k \\
\text{fact}'(k, n) &= \text{fact}'(k \times n, n - 1)
\end{align*}
\]

which is a tail-recursive version of \text{fact}.
4 The monoid

According to [Gil76], a binary algebra \( S = \langle S; \oplus \rangle \) is a monoid iff \( \oplus \) is associative, i.e. for all \( a, b, c \in S \):

\[
a \oplus (b \oplus c) = (a \oplus b) \oplus c
\]

and an element \( 1_\oplus \in S \) exists, such that, for all \( a \in S \),

\[
1_\oplus \oplus a = a
\]

\[
a \oplus 1_\oplus = a
\]

In the examples, we encountered two monoids: \( \langle \text{List}; \rhd \rangle \), and \( \langle \text{Nat}; \times \rangle \), with unit elements \([\ ]\) and \(1\), respectively.

The functions in the examples are of the form:

\[
f(x) = \begin{cases} 
    \text{if } T(x) \\
    \text{then } 1_\oplus \end{cases} \\
\text{else } f(K(x)) \oplus H(x) \\
\text{fi}
\]

so they compute

\[
1_\oplus \oplus (x_1 \oplus (x_2 \oplus (\ldots)\ldots))
\]

with \( x_i = K^i(x) \) (\( i = 1 \ldots n \) for some \( n \) such that \( T(K^n(x))) \), which is equal to (by associativity)

\[
(\ldots (1_\oplus \oplus x_1 \oplus x_2) \oplus \ldots)
\]

so, an equivalent definition is

\[
f(x) = f'(x, 1_\oplus)
\]

\[
f'(x, y) = \begin{cases} 
    \text{if } T(x) \\
    \text{then } y \\
    \text{else } f'(K(x), H(x) \oplus y) \\
    \text{fi}
\end{cases}
\]

The crucial step in a unfold-fold style derivation is the definition of \( f' \) in terms of \( f \):

\[
f'(x, y) = f(x) \oplus y
\]

(The generalisation step). This transformation was first presented in a comparable form by Darlington and Burstall [Dar72, DB76].

\footnote{This is not essential: when \( T(x), f(x) \) may also be \( G(x) \) for the derivation to be possible. In that case, \( f'(x, y) \) will be \( G(x) \oplus y \) when \( T(x) \).}
5 A taxonomy

In the literature, roughly speaking three strategies that allow the above derivation can be discerned.

The simple accumulation strategy allows only the derivation of accumulator versions of functions, as in the above examples.

*Embedding of the domain* (a parameter is added to a function) is the underlying strategy for accumulation, but also for other strategies like *finite differencing* [PK82].

*Continuation-based transformation strategies* also allow the derivation of programs with functions as parameters.

All three strategies rely on the generalisation of a constant to a parameter; the constant may not always be visible, however.

6 On the equivalence of certain computations

In the early sixties, the advent of ALGOL, and McCarthy laying the foundation for a mathematical theory of computation [McC60], paved the way for the formal study of programs and algorithms.

The first paper I found that deals with accumulation is [Coo66]. There, one can already find the classical examples: factorial and fast reverse. By generalising those functions and the properties that guarantee the equivalence of their different versions, Cooper derives a general theorem for proving the equivalence of certain computations.

Cooper also gives the first formal treatment of *inverting the order of evaluation* [Boi92] for the factorial function.

7 Some descriptions by R. Bird and others

In [Bir84a], the use of the accumulation strategy in transformational programming is described. Bird considers *parameter accumulation* to be one of the most important things to be taught in functional programming. In his notation, the conditions for applicability of parameter accumulation are given as follows:

- Given a specification of the form

\[ \text{spec } x = f(H \ x) \]
where $H x$ is a “large” object, defined by (when $x$ is a list)

$$
H[ ] = \ldots \\
H(a; x) = h a (H x)
$$

and a function $f'$ exists such that

$$
f x = f' c x
$$

for some constant $c$;

- and the following condition holds:

$$
(f' s) \circ (h \ a) = (h' a) \circ (f'(g a s))
$$

for suitable functions $g$ and $h'$,

- then $spec$ can be generalised to a function

$$
\text{spec}' \ s \ x = f' \ s \ (H \ x),
$$

- and we get

$$
\text{spec} \ x = \text{spec}' \ c \ x \\
\text{spec}' \ s(a; x) = h' a(\text{spec}(g a s)x).
$$

Lately, R. Bird has been developing a formalism with L. Meertens and others, which is colloquially called Squiggol [STO89, Bir87, Mee86]. In Squiggol, the term “accumulation” has taken on a slightly different meaning. The accumulation of an operator $\oplus$ over a list $L$, $\#_e L$, returns the list containing all intermediate results of a directed reduction, i.e.

$$
\#_e [a_1, a_2, \ldots, a_n] = [e, e \oplus a_1, \ldots, (e \oplus a_1) \oplus a_2 \ldots] \oplus a_n\\
\#_e [a_1, a_2, \ldots, a_n] = [a_1 \oplus (a_2 \oplus \ldots (a_n \oplus e)\ldots), \ldots, a_n \oplus e, e]
$$

This bears some relation to the accumulation described in [Bir84a], in the sense that the “final result” is built up iteratively. The related optimising transformation is comparable to the relation between reductions on one hand and directed reductions and accumulations on the other hand (cf. [Bir87]). One could also argue that accumulations in the traditional sense are less important in Squiggol because of the use
of the (undirected) reduction operator which makes associativity and, often, unit elements, implicit. Informally, the reduction operator can be characterised by:

\[ \oplus [a_1, a_2, \ldots, a_n] = a_1 \oplus a_2 \oplus \ldots \oplus a_n \]

where the absence of brackets indicates associativity of \( \oplus \).

The accumulation strategy as described in this paper is called (one of the applications of) tupling in more recent Squiggol terminology. Meertens recently proposed a class of functions called paramorphisms [Mee92]. Using a form of tupling, paramorphisms can be transformed into homomorphisms. This construction is also used by Harrison [Har91, fold-axiom R3].

Recent work by Meijer, Fokkinga and Paterson [MFP91] introduces a still wider class of functions, called hylomorphisms. Erik Meijer has shown that accumulation can be easily handled in this context [Mei91].

8 Descriptions in CIP-L

In the Munich CIP-project [BBB85], accumulation is mainly viewed as one of the possibilities for transforming a linear recursive function into tail-recursive form. This can clearly be seen in the textbooks [BW82, Par90].

Bauer and Wössner describe the accumulation with the monoid conditions as a special instance of accumulation with an associative dual, i.e., \( \oplus \) is of type \( P \times Q \rightarrow P \), and an operator \( \otimes \) (the associative dual of \( \oplus \)) of type \( Q \times Q \rightarrow Q \) exists, such that

\[ (a \oplus x) \oplus y = a \oplus (x \otimes y). \]

This condition, with the existence of a right identity of \( \oplus \), is also sufficient for accumulation (by definition of \( f'(x, y) = f(x) \oplus y \)). See also [BT90] for a more extensive treatment of associative duals and their properties.

In [Par90], accumulation is described as follows:

**Simplification of linear recursion I**

\[
\text{funct } f = (m \; x)m:
\begin{align*}
\text{if } B(x) \text{ then } H(x) \text{ else } & p(K(x), f(K'(x))) \text{ fi} \\
\text{ASSOC}(m, p) & \\
\text{NEUTRAL}(m, p, E)
\end{align*}
\]
\begin{align*}
\text{funct } f &= (m \, x)m:
\quad f'(E, \, x) \text{ where}
\text{funct } f' &= (m \, y, \, m \, x)m:
\quad \text{if } B(x) \text{ then } p(y, \, H(x)) \text{ else } f'(p(y, \, K(x)), \, K'(x)) \fi
\end{align*}

where \texttt{ASSOC} and \texttt{NEUTRAL} denote degenerated algebraic types which state that the function \( p \) is associative on type \( m \) and that \( E \) is a neutral element (identity) of the function \( p \) on type \( m \), respectively. These are the monoid conditions, as expected.

Of course, this transformation rule can also be viewed as the summary of a schematic transformational development, starting from the input scheme of the rule:

**Embedding:**

define new function \( f' \) by
\[
  f'(m \, y, \, m \, x) = p(y, \, f(x))
\]

**Development:**

**Focus:** function \( f' \)

**Goal:** definition of \( f' \) independent of \( f \)

**Strategy:** unfold-fold

**Transformations:**

step 1: unfold \( f \)
\[
p(y, \, f(\text{if } B(x) \text{ then } H(x) \text{ else } p(K(x), \, f(K'(x))) \fi))
\]

step 2: distribution of function call over conditional
\[
\text{if } B(x) \text{ then } p(y, \, H(x)) \text{ else } p(y, \, p(K(x), \, f(K'(x))))) \fi
\]

step 3: associativity of \( p \)
\[
\text{if } B(x) \text{ then } p(y, \, H(x)) \text{ else } p( \, p(y, \, K(x)), \, f(K'(x))) \fi
\]

step 4: fold \( f' \)
\[
\text{if } B(x) \text{ then } p(y, \, H(x)) \text{ else } f'( \, p(y, \, K(x)), \, K'(x)) \fi
\]

The condition \texttt{NEUTRAL}(\( m \), \( p \), \( E \)) comes in when \( f \) is subsequently defined in terms of \( f' \).

Note that the \textit{embedding} (of the domain) is the crucial step in the derivation. Generally, it is well recognised in the CIP-methodology that embedding (or \textit{generalisation}) is a very important method in program transformation.
9 The Edinburgh generalisation strategy

Among the first to recognise the importance of generalisation in program transformations was the theorem proving group in Edinburgh [BM75, Aub75]. The generalisation strategy is based on the well-known heuristic that, in order to prove a theorem, one should sometimes prove a more general theorem. In functional programming, this translates to defining more general functions, e.g., functions with extra arguments. The choice between possible generalisations is based on mismatch information. Obviously one of the applications of this strategy is accumulation (called generalisation of output terms in [Aub75]), where the unfold-fold derivation starts with a more general function. In [Aub75], one can also find a derivation of the fast reverse.

10 Accumulation by efficient representation of continuations

In the area of program development, one of the most important issues is reducing the number of eurekas or rabbis [Bro89, vdW89]. Steps that appear to be inventive need a simple motivation rather than “a little foresight”.

In my opinion, one of the best motivations given for accumulation is the one by Wand [Wan80]. In his paper, he describes a simple strategy:

given a function $f$ with argument $x$, define a new function $f'$ by

$$f'(\gamma, x) = \gamma(f(x)).$$

Thus, $f(x) = f'(\lambda y. y, x)$. $\gamma$ is a continuation function. Then derive a definition of $f'$ independent of $f$. In many cases, a more efficient representation of the continuation function suggests itself (at this point, there is still a (small?) eureka). This is a generalisation; an (invisible) occurrence of the identity function (a constant) is generalised to an arbitrary function.

Accumulator versions of functions can be derived as follows:

given a function of the form

$$f(x) = \begin{cases} T(x) & \text{if } T(x) \\ H(x) & \text{else} \end{cases}$$

define

$$f'(\gamma, x) = \gamma(f(x)).$$
then

\[ f'(\gamma, x) = \{\text{unfold } f\} \quad \gamma(\begin{cases} \text{if } T(x) \text{ then } 1 & \\
\text{else } f(K(x)) \oplus H(x) & \end{cases} \phi) \]

\[ = \{\text{distributivity}\} \quad \text{if } T(x) \text{ then } \gamma(1) \\
\text{else } \gamma(f(K(x)) \oplus H(x)) & \phi \]

\[ = \{\text{abstract}\} \quad \text{if } T(x) \text{ then } \gamma(1) \\
\text{else } (\lambda y. \gamma(y \oplus H(x)))(f(K(x))) & \phi \]

\[ = \{\text{fold } f\}' \quad \text{if } T(x) \text{ then } \gamma(1) \\
\text{else } f'(\gamma \circ (\lambda y. y \oplus H(x)), K(x)) & \phi. \]

Now it can be shown that every function \( \gamma \) is of the form \( \lambda y. y \oplus z \) for some \( z \):

\[ \lambda y. y = \{\text{trivial}\} \quad \lambda y. y \oplus 1; \]

\[ \gamma \circ (\lambda y. y \oplus H(x)) = \{\gamma \text{ is } \lambda y. y \oplus z\} \quad \lambda y. (y \oplus H(x)) \oplus z \]

\[ = \{\text{associativity } \oplus\} \quad \lambda y. y \oplus (H(x) \oplus z) \square \]

Thus, \( \gamma = \lambda y. y \oplus z \) can be represented by just \( z \). Note that \( \gamma(1) \) simplifies to \( z \).

This gives the well-known accumulator version of \( f \):

\[ f(x) = f''(1, x) \]

\[ f''(z, x) = \begin{cases} z & \text{if } T(x) \\
\text{else } f''(H(x) \oplus z, K(x)) & \end{cases} \phi. \]

It is important to note that there is not much operational difference between the original \( f \) and \( f' \). In \( f' \) the continuation function represents the actions that have to be taken after the recursion has terminated, while this is left implicit in \( f \). Generally, transformations that change or simplify recursion can be applied by first making explicit part of the recursion mechanism and then proving properties of these representations of recursion (e.g., stacks, cf. the discussion in [Boi92]).

11 Pettorossi’s higher order generalisation

Pettorossi and Skowron’s lambda abstraction or higher order generalisation strategy [PS87, Pet87, PS90] takes the ideas from [Wan80] even further. As in the Edinburgh generalisation strategy, mismatch information is used to motivate the introduction of functions with functions as extra arguments. Their goal usually is the derivation of tail-recursive functions, as in the CIP methodology.
The application of lambda abstraction to the standard accumulation examples leads to curried versions of the transformed functions.

Often, the lambda abstraction strategy is used in conjunction with the tupling strategy [Pet84]. Apart from accumulation, lambda abstraction can also be used to derive programs with continuations, e.g. for deriving complicated programs for traversing data structures in one pass as in [Bir84b]. Related techniques are described in [Hug82, Joh85, Tak87].

12 Hughes’ novel implementation of lists

In [Hug86], the idea is presented that a list $x$ could be represented by

\[ \text{rep}(x) = \lambda y. x \mathbin{\#} y, \]

with corresponding abstraction function

\[ \text{abs}(x) = x[] . \]

It is shown how this allows a straightforward derivation of the “fast reverse” (cf. section 2). Also, concatenation on the abstract level is represented by function composition on the concrete level, since

\[
\text{rep}(x \mathbin{\#} y) = \begin{cases} 
\text{definition rep} & \lambda z. (x \mathbin{\#} y) \mathbin{\#} z \\
\text{associativity \#} & \lambda z. x \mathbin{\#} (y \mathbin{\#} z) \\
\text{definition rep} & \lambda z. \text{rep}(x)(\text{rep}(y)z) \\
\text{\text{-conversion}} & \text{rep}(x) \circ \text{rep}(y).
\end{cases}
\]

At a first glance, this seems a revolutionary idea. I will demonstrate that, when one generalises this idea to arbitrary data types with an arbitrary binary operation, it appears that, again, the monoid conditions should hold for this kind of implementations of data types to be useful.

Given a data type $D$ with function $\oplus$ of type $D \times D \to D$ (or $D \to D \to D$), let

\[
\begin{align*}
\text{rep}(d) &= (d \oplus) \\
\text{abs}(x) &= x.1_{\oplus}.
\end{align*}
\]

This is always a valid implementation of $D$, since

\[ \text{abs}(\text{rep } d) = \text{abs}(d \oplus) = d \oplus 1_{\oplus} = d. \]
The equation
\[ \text{rep}(\text{abs } x) = x \]
only holds for \( x \) (a function from \( D \to D \)) that can be written as \( \lambda y.x' + y \):

\[
\text{rep}(\text{abs } x) = \{\text{def. rep} \} \lambda z.(\text{abs } x) + z \\
= \{\text{def. abs} \} \lambda z.(x \ 1_\oplus) + z \\
= \{x \text{ is } \lambda y.x' + y \} \lambda z.x' + z \\
= \{\alpha\text{-conversion} \} \ x
\]

When does this constitute a *useful* implementation of \( D \)? The most important aspect is whether \( \oplus \) can be efficiently implemented in the new representation:

\[
\text{rep}(x + y) = \{\text{def. rep} \} \lambda z.(x + y) + z \\
= \{\text{associativity of } \oplus \text{ necessary} \} \lambda z.x + (y + z) \\
= \{\text{def. rep} \} \lambda z.(\text{rep } x)(\text{rep } y)z \\
= \{\eta\text{-conversion} \} \ (\text{rep } x) \circ (\text{rep } y).
\]

For other operators to be efficiently implementable, certain distributive properties should hold.

So, for implementation of objects \( d \) of a data type \( D \) by a function \( d \oplus \), again the monoid properties should hold. An identity of \( \oplus \) should exist to allow an abstraction function, and \( \oplus \) should be associative for its implementation to be simple. Thus, it is not surprising that this technique led to a derivation of the fast reverse.

### 13 Some more papers

There are many more papers discussing accumulation, generalisation and related techniques.

Burstall and Feather’s early survey paper [BF78] gives the familiar factorial example as an example for generalisation, and more references to papers from the early days of transformational programming.

The classic papers by Strong and others [AS78, Str71, WS73] are mostly concerned with recursion elimination, i.e. with conditions that allow more than the trivial transition from tail recursion to iteration. The associativity conditions are among the most obvious of those.

The paper by Huet and Lang [HL78] uses *recursion removal* as an example of program transformations expressed with second-order patterns. (These second-order
patterns are comparable to the transformation rules in the CIP approach. These examples (mainly taken from Darlington's thesis [Dar72]) include accumulation and accumulation with associative dual.

Manna and Waldinger [MW] also mention generalisation as an important technique in program synthesis, and show a derivation of the fast reverse.

Arsac and Kodratoff [AK82] give a method for finding generalisations that lead to tail-recursive functions. Their examples involve more complicated properties than just associativity; the existence of identities of functions is generally assumed (cf. the adjunction of ‘fictitious’ identity elements in Squiggol [Bir87]). They also briefly consider ‘generalisation using second-order ideas’, where a functional variable is introduced. By explicitly introducing the stacks that are necessary in the evaluation of non-tail-recursive functions, they derive iterative versions of arbitrary recursive functions.

Gabriel [Gab91] calls the accumulation rule *parentheses movement*¹, and uses it as an example of the DEVA meta-calculus. The most impressive aspect of this paper is the \LaTeX\-representation of the (really two-dimensional) DEVA fragments, which has been automatically generated by the DEVA system.

Acknowledgement

I thank Alberto Pettorossi for introducing me to many of the papers mentioned above and for interesting discussions on this and other subjects during my stay at IASI in Rome. Helmut Partsch also provided numerous references and useful comments on this paper.

References


¹The German version of Bauer and Wössner’s book has “Klammerverschiebung”.

13


