FAIRNESS IN CONTEXT-FREE GRAMMARS UNDER CANONICAL DERIVATIONS

by

S. Porat and N. Francez

Technical Report #340
October 1984
FAIRNESS IN CONTEXT-FREE GRAMMARS UNDER CANONICAL DERIVATIONS

by
Sara Porat, Nissim Francez
Computer Science Department
The Technion, Haifa 32000, ISRAEL

Summary
In [PFMZ 82] the notion of fair derivations in context-free grammars was introduced and studied. The main result there is a characterization of fairly terminating grammars as non-variable-doubling. In this paper we show that the same characterization is valid under canonical derivations in which the next variable to be expanded is deterministically chosen, leaving nondeterminism only to the decision as to which rule to apply. Two families of canonical derivations are introduced and studied: 1) Spinal derivations and 2) Layered derivations.
1. Introduction

In [PFMZ 82] the concept of fair derivations in context free grammars was introduced in order to study the effects of fairness assumptions in a more abstract context than the usual context of nondeterministic and concurrent programming. The main result of that paper is a characterization of fairly terminating CF grammars as equivalent to variable doubling (or expansive) CF grammars. This characterization establishes the decidability question for fair termination of CF grammars, in contrast to the highly undecidable nature of fair termination in high level nondeterministic programming languages [H 64].

While obtaining that characterization it was noted that CF grammars actually contain two contexts in which nondeterministic choices are applied in the next step in some derivation:

1. The choice of the variable in the sentential form to be replaced in the next step in the derivation.
2. The choice of the production rule to be applied to the chosen variable.

The way variable doubling was used to obtain infinite fair derivations was by using one copy of such a variable to generate future occurrences of that variable while the second copy was used to apply, in some fixed order, all the other productions of the variables occurring infinitely often in the derivation.

The motivation for part of the study reported here was based on a dissatisfaction from the way fair behaviors reflected in that context: it is achieved by applying rules of the same variable in independent subderivations. Thus, in tracing the infinite chain of descendants of a specific occurrence of a variable, it need not be the case that indeed all its rules are applied in that chain, as should be the case in some natural conception of fairness in such derivations.

In this paper we suggest a more restrictive notion of enabledness of a production rule by eliminating the first context of nondeterminism: we fix deterministically the way the next variable to be replaced is chosen, leaving nondeterminism only in the choice of the next rule to be applied. By this restriction, less derivations are considered.

We deal with two families of (canonical) infinite derivations:

1. Spinal derivations: these are (infinite) derivations in the derivation trees of which there is only one infinite path. Known examples of such derivations are the leftmost and the rightmost derivations. In such derivations, the descendant variable occurrences of any given variable occurrence in a form are replaced before any “sibling” occurrence is replaced. Such derivations seem to avoid the dissatisfaction mentioned above. As a representative of this family we shall consider the leftmost derivations.

2. Layered derivations: these are (infinite) derivations in the derivation trees of which the leaves are always labeled by terminals. The variable occurrences are replaced in such a way that no variable occurrence is left unexpanded for ever. We consider, as a representative of this family, derivations where replacements are performed in an order dictated by the distance from the root; for equidistant variable occurrences a left-to-right order is imposed.

The main result we prove is that the same characterization of a fairly terminating CF grammar is valid. Thus, variable doubling seems to have a more fundamental role.

Another study of fair termination in the context of formal languages, inspired by [PFMZ 82], may be found in [RA 64], where fair termination of EOL systems is studied.

2. C-Fair derivations in CF grammars

In the sequel we use standard notation for CF grammars and languages [HU 79]. The results are comprehensible without prior familiarity with [PFMZ 82], though their importance might be better appreciated by readers familiar with the previous treatment.
Recall that in the derivations considered in this section the variable occurrence to be replaced is deterministically chosen under a choice-strategy $C$. These derivations are referred to as $C$-derivations. A $C$-derivation starting from a sentential form $\alpha$ to a sentential form $\beta$ is denoted by $\alpha \xrightarrow{C} \beta$. Let $G = (V,T,P,S)$ be a context free grammar. We assume that all the context free grammars referred to have no useless variables.

Definitions:
1. A production rule $A \rightarrow \alpha \in P$ is $C$-enabled in a sentential form $\beta$ iff $A$ is the variable whose occurrence is chosen as next under the strategy $C$.
2. A $C$-derivation $d$ is $C$-fair iff it is finite or it is infinite and every rule that is infinitely-often $C$-enabled along $d$ is also infinitely-often applied along $d$.
3. A CF grammar $G$ is $C$-fairly terminating iff all its $C$-fair derivations are finite.

Remark: For linear CF grammars, these definitions coincide with the fair derivations as defined in [PFMZ 82] as sentential forms contain at most one variable occurrence and have no nondeterminism in variable occurrence choices.

Definition: A CF grammar is variable doubling (expansive) iff there is a variable $A \in V$ such that
$A \rightarrow^{*} \alpha_1 A \alpha_2 A \alpha_3$
for some $\alpha_1, \alpha_2, \alpha_3 \in (V \cup T)^*$.

Example: Consider the following grammar $G_1$ (table 1). This grammar is variable doubling ($A$ doubling), as is seen from the derivation tree presented in figure 1.

Remark: note that in contrast to the situation in the case of arbitrary derivations, there need not exist a sentential form (under a strategy $C$) derived from $A$ in which $A$ occurs twice. Our main concern in the sequel is to overcome this obstacle. we show in the sequel that variable doubling is a sufficient condition for the existence of an infinite $C$-fair derivation for the strategies considered. Note that the construction presented in [PFMZ 82] is not applicable here, since it uses arbitrary

$G_1 = \{S, A, B, C, \{a, b, c\}, P, S\}$

where $P$ contains the rules:
$S \rightarrow A$
$A \rightarrow BSC | b$
$B \rightarrow aA$
$C \rightarrow c$

Table 1: The grammar $G_1$
Figure 1: derivation tree of A-doubling in $G_1$

derivations, while we have to show the existance of fair $C$-derivations.

2.1 Fair spinal derivations

The strategy $C$ is now that of leftmost derivations. All the results in this subsection apply to arbitrary spinal derivations and the leftmost rule is chosen for notational convenience.

**Definition:** A variable occurrence $A$ is next under the leftmost strategy ($L$) in a sentential form $\beta$ iff $\beta = wA\gamma$ for $w \in T^*, \gamma \in (V \cup T)^*$.

**Remark:** According to this definition of a strategy, a rule is enabled on a sentential form whenever its LHS variable has an occurrence the leftmost variable in the form. We now use $L$ instead of the generic $C$ in the definitions above.

**Example:** We present an example of an infinite $L$-unfair derivation in $G_1$.

$$S \rightarrow A \rightarrow BSC \rightarrow aASC \rightarrow aBSCSC \rightarrow a^iA(SC)^i \rightarrow a^iB(SC)^i \ldots \text{ for } i \geq 2$$

This infinite derivation is $L$-unfair since $A$ is infinitely often the next (to be replaced), but the rule $A \rightarrow b$ is never applied.

Note that there are no $a_1, a_2, a_3 \in (V \cup T)^*$ such that

$$A \rightarrow L a_1Aa_2Aa_3$$

**Theorem:** (characterization of $L$-fair termination) A CF grammar is $L$-fairly terminating iff it is not variable doubling.

**Proof:**

(if) We show that if a grammar $G$ has an infinite $L$-fair derivation then it doubles some variable. Note that such a derivation is not necessarily a fair derivation according to the definition in [PFMZ 82], as variable occurrences which are not leftmost need not be expanded in an $L$-fair derivation, but are enabled according to that derivation. Thus, the previous proof cannot be used. We show this by an example.

**Example:** Consider the infinite $L$-fair derivation, referred to as $d_f$, as follows:
$S \rightarrow A \rightarrow BSC \rightarrow aASC \rightarrow \cdots$

(i.e. repeating the same sequence of rule applications indefinitely).

As $C$ appears infinitely often along $d_1$ but the (unique) C-rule is never applied, this derivation is unfair according to the definition in [PFMZ 82].

The general form of a derivation tree of an infinite $L$-derivation is as shown in figure 2. It contains exactly one infinite path referred to as the spine. The subtrees to the left of the spine are all finite. To the right of the spine are only leaves labeled either by terminals or by variables.

Claim: there exists a variable $A \in V$ that has infinitely-many occurrences along the spine and there is an $A$-rule applied only finitely often along the spine.

To show the claim, partition $V$ into

$V = V^1 \cup \cdots \cup V^n$

where

$A \in V^i \leftrightarrow i = \min\{j \mid A \rightarrow^j w \in T^\ast\}.$

Clearly, as there are no useless variables, every variable is covered by this partition.

Since the spine is infinite, obviously there is a variable $A$ occurring infinitely often along it. If one of these infinitely-repeating variables is a member of $V^1$, the claim immediately follows (as it has a rule with a terminal R.H.S that could not have been applied along the spine).

Otherwise, let $i_c > 1$ be the minimal index such that an infinitely-repeating variable $A$

![Figure 2: an infinite L-derivation tree](image-url)
belongs to $V^o$. Suppose, by way of contradiction, that all the $A$-rules have been applied infinitely often along the spine. Necessarily, $A$ has a rule in which all variables at the RHS belong to $V^j$'s with $j<i_0$, contradicting the minimality of $i_0$. This proves the claim.

Let $A$ be the variable existing by the claim, and $A \rightarrow \alpha$ be the corresponding rule. Consider the subtree $T_A$ of the given $L$-fair derivation, the root of which is the first occurrence of $A$ along the spine, from which $A \rightarrow \alpha$ is not applied anymore. Such an occurrence of $A$ exists by the assumption that $A \rightarrow \alpha$ is applied only finitely often along the spine. The situation is shown in figure 3.

As the whole derivation is given to be $L$-fair and $A$ occurs infinitely often from that point onwards, there exists a finite subtree to the left of the spine where the rule $A \rightarrow \alpha$ is applied. Thus, the subtree $T_A$ has two nodes labeled $A$: one of them on the
spine, the other on the left finite subtree. We show an example of a fair spinal derivation tree in figure 4.

Along the spine $A \in V$ appears infinitely often but the rule $A \rightarrow b$ is applied only outside the spine, in a left finite subtree. Therefore, $A$ doubles itself in $G_1$.

Thus, the grammar has a variable doubling itself, establishing this direction of the theorem.

(only if)

We have to show that if the grammar has a variable doubling itself then it has an infinite $L$-fair derivation.

Let $A \in V$, $a_1,a_2,a_3 \in (V \cup T)^*$ such that

$$A \rightarrow a_1 A a_2 A a_3.$$

The idea is to construct an infinite $L$-fair derivation in such a way that $A$ is "responsible" for the fair application of rules. We design the derivation so that $A$ is encountered alternately, once in a left finite subtree and the second time down the spine. This is possible due to the $A$-doubling property.

Formally, let

$$A \rightarrow w_1 B \cdots w_m B^m w_{m+1},$$

where

$$w_i \in T^*, 1 \leq i \leq m+1$$

and

$$B^i \in V, 1 \leq j \leq m$$

be the first $A$-rule applied along the derivation in which $A$ doubles itself. Also, let

$$B^i, B^j, i < j$$

be the variables such that

---

Figure 4: a spinal derivation tree of an $L$-fair derivation in $G_1$
We now construct two $L$-subderivations that repeat themselves infinitely often in the $L$-fair derivation. The first one, denoted by $d_1$, starts in $A$ and ends in a form where $A$ is again the next variable to be replaced. In this form there is another variable that eventually derives another occurrence of $A$ (the one along the spine).

$$A \rightarrow w_1 B^t \cdots w_i B^t \cdots w_j B^t \cdots w_m B^m w_{m+1} \rightarrow^L t_1$$

$$t_1u_1 A \beta_1 \cdots w_m B^m w_{m+1} \rightarrow^L$$

where

$$t_1 \in \mathcal{T}^*.$$

The second $L$-subderivation, denoted by $d_2$, ends with a form containing the next spinal occurrence of $A$.

$$\beta_1 w_{i+1} \cdots w_j B^t \cdots w_m B^m w_{m+1} \rightarrow^L t_2$$

$$t_2u_2 A \beta_2 \cdots w_m B^m w_{m+1} \rightarrow^L$$

where

$$t_2 \in \mathcal{T}^*.$$

Also, for every variable $D$ derivable from $A$ and for every $D$-rule $D \rightarrow \gamma$, define an $L$-subderivation, denoted by $d_{D, \gamma}$. It starts in $A$ and contains an application of the rule $D \rightarrow \gamma$. It ends with a terminal word.

$$A \rightarrow w_D \alpha_D \rightarrow w_D \gamma \alpha_D \rightarrow^L w_D$$

for some

$$w_D, \alpha_D \in \mathcal{T}^*, \alpha_D \in (V \cup \mathcal{T})^*.$$

We now impose the ordering in which the rules will be applied as a round robin, ensuring fairness. Let $D_1, \ldots, D_k$ be all the variables derivable from $A$. For every $D_i$, let

$$\{D_i \rightarrow \gamma_1, \ldots, D_i \rightarrow \gamma_{k_i}\}$$

be all the $D_i$-rules.

We combine all these subderivations into a section of an $L$-derivation, denoted by $d_A$ and referred to as a full cycle, that by repeating itself will generate the required infinite $L$-fair derivation.

We start by applying the subderivation $d_1$. Afterwards $d_{D_1, \gamma_1}$; afterwards $d_{D_1, \gamma_2}$ followed by $d_2$ and then $d_{D_2, \gamma_1}$, etc., until all the $D_i$-rules have been applied in turn. At this stage we start a similar section, this time "being fair" towards $D_2$ and so on until $D_k$ has been covered. The structure of $d_A$ is shown in figure 5.
We give also an example of a full cycle in $G_i$ in figure 6.

Finally, by prefixing to an indefinite repetition of a full cycle a section that generates $A$ from $S$, we obtain the required infinite $L$-fair derivation.
2.2 Fair layered derivations

In this section we consider another strategy for determinizing the choice of the next variable and its effect on fairness. Recall that in the layered strategy, denoted by LA, the variable occurrences are considered in layers, where a layer consists of all the variables with the same distance from the root in a derivation tree. Within a layer the variables are chosen in a left-to-right order.

To express formally this strategy, we associate with each variable occurrence in a sentential form a natural number, its distance from the root.

**Definition:** A variable occurrence $A$ is next in a form $\alpha$ under the strategy LA iff

$$\beta = \gamma A \delta, \quad \gamma, \delta \in (V \cup T)^*$$

and there exist a natural number $i$ such that the distance of the occurrence of $A$ to the right of $\gamma$ is $i$, and the distances of all the variable-occurrences in $\gamma$ are $i+1$ and these in $\delta$ are (still) at distance $i$. 

![Figure 6: A full cycle in $G_i$](image-url)
The general form of an LA-derivation tree is shown in figure 7. In the figure, a small square denotes a terminal-labeled node while a small circle denotes a variable-labeled node. The distances of the nodes are also indicated.

**Theorem: (characterization of LA-fair termination)**

A CF grammar is **LA-fairly terminating** iff it is not variable doubling.

**Proof:**

(if) This direction is immediately following from the characterization theorem in [PFMZ 82], as every infinite LA-fair derivation is also fair under the definition there.

(only if)

Suppose the given grammar doubles the variable $A$. We first describe a section of an LA-derivation starting with $A$ and guaranteeing that for any variable $B$ derivable from $A$ all the $B$ rules are used. This can be done since the variable-doubling property ensures two occurrences of $A$ (though not necessarily at the same layer); the left one is used for fairly expanding all the variable occurrences while the right one allows another expansion of $A$. The details of the formal construction are similar to the spinal case and all one has to check is that they can be carried out in using LA derivations. We omit the details.

As before, this section is repeated infinitely often prefixed with another section generating a sentential form containing a (first) $A$. Here one has to take care that the form-portions appearing to the left of $A$ and to its right are expanded in such a way to produce finite subtrees (with terminal leaves). This is possible by the assumption of absence of useless variables.

![Figure 7: an LA-derivation tree](image-url)
References


