A short proof that O_2 is an MCFL

Mark-Jan Nederhof
School of Computer Science
University of St Andrews, UK

Abstract

We present a new proof that $\mathbf{O}_2$ is a multiple context-free language. It contrasts with a recent proof by Salvati (2015) in its avoidance of concepts that seem specific to two-dimensional geometry, such as the complex exponential function. Our simple proof creates realistic prospects of widening the results to higher dimensions. This finding is of central importance to the relation between extreme free word order and classes of grammars used to describe the syntax of natural language.

1 Introduction

The alphabet of the *MIX language* has three symbols, a, b and c. A string is in the language if and only if the number of a's, the number of b's, and the number of c's are all the same. A different way of defining the MIX language is as permutation closure of the regular language $(abc)^*$, as noted by Bach (1981); see also Pullum (1983).

If a, b and c represent, say, a transitive verb and its subject and its object, then a string in MIX represents a sentence with any number of triples of these constituents, in a hypothetical language with extreme free word order. This is admittedly rather unlike any actual natural language. Joshi (1985) argued that because of this, grammatical formalisms for describing natural languages should *not* be capable of generating MIX. He also conjectured that MIX was beyond the generative capacity of one particular formalism, namely the tree adjoining grammars. Several decades passed before Kanazawa and Salvati (2012) finally proved this conjecture.

MIX has been studied in the context of several other formalisms. Joshi et al. (1991) showed that MIX is generated by a generalization of tree adjoining grammars that decouples local domination for linear precedence. Boullier (1999) showed that MIX is generated by a range concatenation grammar. Negative results were addressed by Sorokin (2014) for well-nested multiple context-free grammars, and by Capelletti and Tamburini (2009) for a class of categorial grammars. The MIX language is also of interest outside of computational linguistics, e.g. in computational group theory (Gilman, 2005).

A considerable advance in the understanding of the MIX language is due to Salvati (2015), who showed that MIX is generated by a multiple context-free grammar (MCFG). The main part of the proof shows that the language $\mathbf{O}_2$ is generated by a MCFG. This language has four symbols, a, $\overline{a}$, b and $\overline{b}$. A string is in the language if and only if the number of a's equals the number of $\overline{a}$'s, and the number of b's equals the number of $\overline{b}$'s. MIX and $\mathbf{O}_2$ are rationally equivalent, which means that if one is generated by a multiple context-free grammar, then so is the other.

The proof by Salvati (2015) is remarkable, in that it is one of the few examples of geometry being used to prove a statement about formal languages. The proof has two related disadvantages however. The first is that a key element of the proof, that of the complex exponential function, is not immediately understood without background in geometry. The second is that this also seems to restrict the proof technique to two dimensions, and there is no obvious avenue to generalize the result to a variant of MIX with four or five symbols. We hope to remedy this by an alternative, self-contained proof that avoids the complex exponential function. The core idea is a straightforward normalization of paths in two dimensions, which allow simple arguments to lead to a proof by contradiction. We also sketch part of a possible proof in three dimensions.

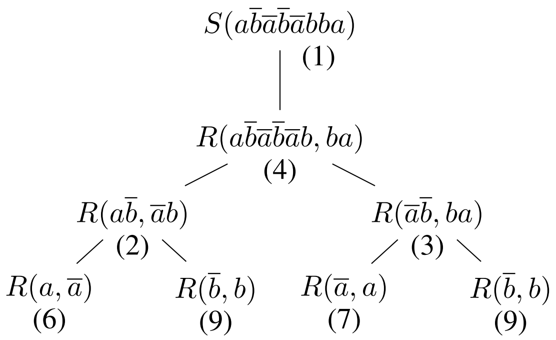

Figure 1: Derivation in G. The numbers indicate the rules that were used.

2 Initial problem

The MCFG G is defined as:

$$
\begin{array}{rcll}
S(xy) & \leftarrow & R(x, y) & (1) \\
R(xp, yq) & \leftarrow & R(x, y)\ R(p, q) & (2) \\
R(xp, qy) & \leftarrow & R(x, y)\ R(p, q) & (3) \\
R(xpy, q) & \leftarrow & R(x, y)\ R(p, q) & (4) \\
R(p, xqy) & \leftarrow & R(x, y)\ R(p, q) & (5) \\
R(a, \bar{a}) & \leftarrow & & (6) \\
R(\bar{a}, a) & \leftarrow & & (7) \\
R(b, \bar{b}) & \leftarrow & & (8) \\
R(\bar{b}, b) & \leftarrow & & (9) \\
R(\varepsilon, \varepsilon) & \leftarrow & & (10)
\end{array}
$$

For the meaning of MCFGs in general, see Seki et al. (1991); for a closely related formalism, see Vijay-Shanker et al. (1987); see Kallmeyer (2010) for an overview of mildly context-sensitive grammar formalisms.

The reader unfamiliar with this literature is encouraged to interpret the rules of the grammar as logical implications, with S and R representing predicates. There is an implicit conjunction between the two occurrences of R in the right-hand side of each of the rules (2) — (5). The symbols x, y, p, q are string-valued variables, with implicit universal quantification that has scope over both left-hand side and right-hand side of a rule. The rules (6) — (10) act as axioms. The symbols a, $\bar{a}$, b, $\bar{b}$ are terminals, and ε denotes the empty string.

We can derive $S(x)$ for certain strings x, and $R(x, y)$ for certain strings x and y. Figure 1 presents an example of a derivation. The language generated by G is the set L of strings x such that $S(x)$ can be derived.

By induction on the depth of derivations, one can show that if $R(x, y)$, for strings x and y, then $xy \in \mathbf{O}_2$. Thereby, if $S(x)$ then $x \in \mathbf{O}_2$, which means $L \subseteq \mathbf{O}_2$. The task ahead is to prove that if $xy \in \mathbf{O}_2$, for some x and y, then $R(x, y)$. From this, $L = \mathbf{O}_2$ then follows.

Let $|x|$ denote the length of string x. For an inductive proof that $xy \in \mathbf{O}_2$ implies $R(x, y)$, the base cases are as follows. If $xy \in \mathbf{O}_2$ and $|x| \leq 1$ and $|y| \leq 1$, then trivially $R(x, y)$ by rules (6) — (10).

Furthermore, if we can prove that $xy \in \mathbf{O}_2$, $x \neq \varepsilon$ and $y \neq \varepsilon$ together imply $R(x, y)$, for $|xy| = m$, for some m, then we may also prove that $x'y' \in \mathbf{O}_2$ on its own implies $R(x', y')$ for $|x'y'| = m$. To see this, consider $m > 0$ and $z \in \mathbf{O}_2$ with $|z| = m$, and write it as $z = xy$ for some $x \neq \varepsilon$ and $y \neq \varepsilon$. If by assumption $R(x, y)$, then together with $R(\varepsilon, \varepsilon)$ and rule (4) or (5) we may derive $R(xy, \varepsilon)$ or $R(\varepsilon, xy)$, respectively. In the light of this, the inductive step merely needs to show that if for some x and y:

- $xy \in \mathbf{O}_2$, $|x| \geq 1$, $|y| \geq 1$ and $|xy| > 2$, and
- $pq \in \mathbf{O}_2$ and $|pq| < |xy|$ imply $R(p, q)$, for all p and q,

then also $R(x, y)$. One easy case is if $x \in \mathbf{O}_2$ (and thereby $y \in \mathbf{O}_2$) because then we can write $x = x_1 x_2$ for some $x_1 \neq \varepsilon$ and $x_2 \neq \varepsilon$. The inductive hypothesis states that $R(x_1, x_2)$ and $R(\varepsilon, y)$, which imply $R(x, y)$ using rule (4).

A second easy case is if x or y has a proper prefix or proper suffix that is in $\mathbf{O}_2$. For example, assume there are $z_1 \neq \varepsilon$ and $z_2 \neq \varepsilon$ such that $x = z_1 z_2$ and $z_1 \in \mathbf{O}_2$. Then we can use the inductive hypothesis on $R(z_1, \varepsilon)$ and $R(z_2, y)$, together with rule (2).

At this time, the reader may wish to read Figure 1 from the root downward. First, $a\bar{b}\bar{a}\bar{b}\bar{a}bba$ is divided into a pair of strings, namely $a\bar{b}\bar{a}\bar{b}\bar{a}b$ and ba. At each branching node in the derivation, a pair of strings is divided into four strings, which are grouped into two pairs of strings, using rules (2) — (5), read from left to right. Rules (2) and (3) divide each left-hand side argument into two parts. Rule (4) divides the first left-hand side argument into three parts, and rule (5) divides the second left-hand side argument into three parts.

What remains to show is that if $z_1 z_2 \in \mathbf{O}_2$, $z_1 \notin \mathbf{O}_2$ and $|z_1 z_2| > 2$, and no proper prefix or proper suffix of z_1 or of z_2 is in $\mathbf{O}_2$, then there is at least one rule that allows us to divide z_1 and z_2 into four strings altogether, say x, y, p, q, of which at least three are non-empty, such that $xy \in \mathbf{O}_2$. This will then permit use of the inductive hypothesis on $R(x, y)$ and on $R(p, q)$.

We can in fact restrict our attention to $z_1' z_2' \in \mathbf{O}_2$, $|z_1' z_2'| > 2$, and no non-empty substring of z_1' or of z_2' is in $\mathbf{O}_2$, which can be justified as follows. Suppose we have z_1 and z_2 as in the previous paragraph, and suppose z_1' and z_2' result from z_1 and z_2 by exhaustively removing all non-empty substrings that are in $\mathbf{O}_2$; note that still $|z_1' z_2'| > 2$. If we can use a rule to divide z_1' and z_2' into x', y', p', q', of which at least three are non-empty, such that $x'y' \in \mathbf{O}_2$, then the same rule can be used to divide z_1 and z_2 into x, y, p, q with the required properties, which can be found from x', y', p', q' simply by reintroducing the removed substrings at corresponding positions.

3 Geometrical view

We may interpret a string x geometrically in two dimensions, as a path consisting of a series of line segments of length 1, starting in some point (i, j). Every symbol in x, from beginning to end, represents the next line segment in that path; an occurrence of a represents a line segment from the previous point (i, j) to the next point $(i + 1, j)$, $\overline{a}$ represents a line segment from (i, j) to $(i - 1, j)$, b represents a line segment from (i, j) to $(i, j+1)$, and $\overline{b}$ represents a line segment from (i, j) to $(i, j - 1)$. If $x \in \mathbf{O}_2$, then the path is *closed*, that is, the starting point and the ending point are the same. If we have two strings x and y such that $xy \in \mathbf{O}_2$ and $x \notin \mathbf{O}_2$, then this translates to two paths, connecting two distinct points, which together form a closed path. This is illustrated in Figure 2.

In the following, we assume a fixed choice of some x and y such that $xy \in \mathbf{O}_2$, $|xy| > 2$, and no non-empty substring of x or of y is in $\mathbf{O}_2$. If we follow the path of x starting in $P[0] = (0, 0)$, then the path ends in some point $P[1] = (i, j)$ such that i is the number of occurrences of a minus the number of occurrences of $\overline{a}$ and j is the number of occurrences of b minus the number of occurrences of $\overline{b}$. This path from $P[0]$ to $P[1]$ will be called $A[0]$. The path of y from $P[1]$ back to $P[0]$ will be called $B[1]$. We generalize this by defining for any integer k: $P[k]$ is the point $(k \cdot i, k \cdot j)$, $A[k]$ is the path of x from $P[k]$ to $P[k + 1]$ and $B[k]$ is the path of y from $P[k]$ to $P[k - 1]$. Where the starting points are irrelevant and only the shapes matter, we talk about paths A and B.

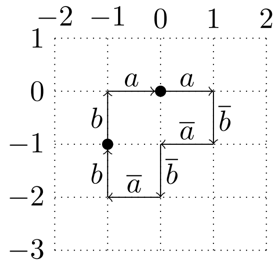

Figure 2: Two strings $x = a\overline{b}\overline{a}\overline{b}\overline{a}b$ and $y = ba$ together represent a closed path, consisting of a path from $(0, 0)$ to $(-1, -1)$ and a path from $(-1, -1)$ to $(0, 0)$.

Let C be a path, which can be either $A[k]$ or $B[k]$ for some k. We write $Q \in C$ to denote that Q is a point on C. Let $Q = (i, j) \in C$, not necessarily with i and j being integers. We define the path-distance $d_C(Q)$ of Q on C to be the length of the path along line segments of C to get from $P[k]$ to Q. In Figure 2, $(0, -1)$ has path-distance 3 on $A[0]$, as the path on $A[0]$ to reach $(0, -1)$ from $P[0] = (0, 0)$ consists of the line segments represented by the prefix $a\overline{b}\overline{a}$ of x. Similarly, $d_{A[0]}((0.5, -1)) = 2.5$.

Let C be a path as above and let points $Q_1, Q_2 \in C$ be such that $d_C(Q_1) \leq d_C(Q_2)$. We define the subpath $D = sub_C(Q_1, Q_2)$ to be such that $Q \in D$ if and only if $Q \in C$ and $d_C(Q_1) \leq Q \leq d_C(Q_2)$, and $d_D(Q) = d_C(Q) - d_C(Q_1)$ for every $Q \in D$. For two points Q_1 and Q_2, the line segment between Q_1 and Q_2 is denoted by $seg(Q_1, Q_2)$.

The task formulated at the end of Section 2 is accomplished if we can show that at least one of the following must hold:

- the angle in $P[0]$ between the beginning of $A[0]$ and that of $B[0]$ is 180° (Figure 3);
- there is a point $Q \notin \{P[0], P[1]\}$ such that $Q \in A[0]$ and $Q \in B[1]$ (Figure 4);
- there is a point $Q \neq P[1]$ such that $Q \in A[0]$, $Q \in A[1]$ and $d_{A[0]}(Q) > d_{A[1]}(Q)$ (Figure 5); or
- there is a point $Q \neq P[0]$ such that $Q \in B[0]$, $Q \in B[1]$ and $d_{B[1]}(Q) > d_{B[0]}(Q)$ (analogous to Figure 5).

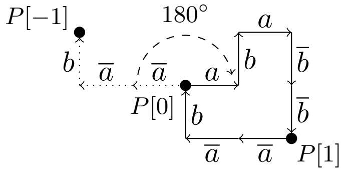

Figure 3: With $x = a\,b\,a\,\overline{b}\,\overline{b}$ and $y = \overline{a}\,\overline{a}\,b$, the beginning of path $A[0]$ and the beginning of (dotted) path $B[0]$ have an 180° angle in $P[0]$, which implies x and y start with complementing symbols (here a and $\overline{a}$; the other possibility is b and $\overline{b}$). By applying rule (2), two smaller closed paths result, one of which consists of these two complementing symbols.

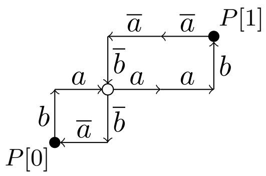

Figure 4: The paths $A[0]$ and $B[1]$ of $x = baaab$ and $y = \overline{a}\,\overline{a}\overline{b}\,\overline{b}\overline{a}$ have point $(1, 1)$ in common. Two smaller closed paths result by applying rule (3).

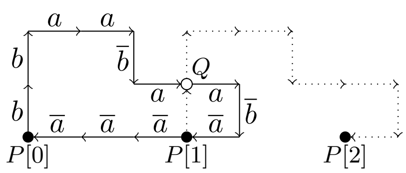

Figure 5: With $x = b\,b\,a\,a\,\overline{b}\,a\,a\,\overline{b}\,\overline{a}$ and $y = \overline{a}\,\overline{a}\,\overline{a}$, the path $A[0]$ and the (dotted) path $A[1]$ have point Q in common, with $d_{A[0]}(Q) = 6 > d_{A[1]}(Q) = 1$. By applying rule (4), two smaller closed paths result, one of which is formed by prefix b of length 1 and suffix $a\overline{b}\,\overline{a}$ of length $|x| - 6 = 3$ of x.

We will do this through a contradiction that results if we assume:

(i) the angle in $P[0]$ between the beginning of $A[0]$ and that of $B[0]$ is *not* 180°;

(ii) $A[0] \cap B[1] = \{P[0], P[1]\}$;

(iii) there is no $Q \in (A[0] \cap A[1]) \setminus \{P[1]\}$ such that $d_{A[0]}(Q) > d_{A[1]}(Q)$; and

(iv) there is no $Q \in (B[0] \cap B[1]) \setminus \{P[0]\}$ such that $d_{B[1]}(Q) > d_{B[0]}(Q)$.

In the below, we will refer to these assumptions as the *four constraints*.

4 Continuous view

Whereas paths A and B were initially formed out of line segments of length 1 between points (i, j) with integers i and j, the proof becomes considerably easier if we allow i and j to be real numbers. The benefit lies in being able to make changes to the paths that preserve the four constraints, to obtain a convenient normal form for A and B. If we can prove a contradiction on the normal form, we will have shown that no A and B can exist that satisfy the four constraints.

We define, for each integer k, the line $\ell[k]$, which is perpendicular to the line through $P[k]$ and $P[k+1]$, and lies exactly half-way between $P[k]$ and $P[k+1]$. Much as before, we write $Q \in \ell[k]$ to denote that Q is a point on line $\ell[k]$. We will consistently draw points $\ldots, P[-1], P[0], P[1], \ldots$ in a straight line from left to right.

Let C be a path, which can be either $A[k']$ or $B[k']$, for some k', and let $Q \in C$. We write $\mathit{from\_right}_C(Q, \ell[k])$ to mean that path C is strictly to the right of $\ell[k]$ just before reaching Q, or formally, there is some $\delta > 0$ such that each $Q' \in C$ with $d_C(Q) - \delta < d_C(Q') < d_C(Q)$ lies strictly to the right of $\ell[k]$. The predicates *from_left*, *to_right*, *to_left* are similarly defined.

Let $Q_1, Q_2 \in C \cap \ell[k]$, for some k, such that $d_C(Q_1) \leq d_C(Q_2)$. We say that C has an *excursion from the right* between Q_1 and Q_2 at $\ell[k]$ if $\mathit{from\_right}_C(Q_1, \ell[k])$ and $\mathit{to\_right}_C(Q_2, \ell[k])$. This is illustrated in Figure 6: the path is strictly to the right of $\ell[k]$ just before reaching Q_1. From there on it may (but need not) cross over to the left of $\ell[k]$. Just after it reaches Q_2, it must again be strictly to the right of $\ell[k]$. The definition of *excursion from the left* is symmetric. Note that excursions may be nested; in Figure 6, $\mathit{sub}_C(Q_1, Q_2)$ has an excursion at $\ell[k]$ from the left below Q_2.

In Figure 6, the pair of points Q_1 and R_1 will be called a *crossing* of $\ell[k]$ from right to left, characterized by $Q_1, R_1 \in \ell[k]$, $\mathit{from\_right}_C(Q_1, \ell[k])$, $\mathit{to\_left}_C(R_1, \ell[k])$ and $\mathit{sub}_C(Q_1, R_1)$ being a line segment. The pair of points R_2 and Q_2 is a crossing of $\ell[k]$ from left to right, where the length of $\mathit{seg}(R_2, Q_2)$ happens to be 0. In much of the following we will simplify the discussion by assuming crossings consist of single points, as in the case of $R_2 = Q_2$. However, existence of crossings con-

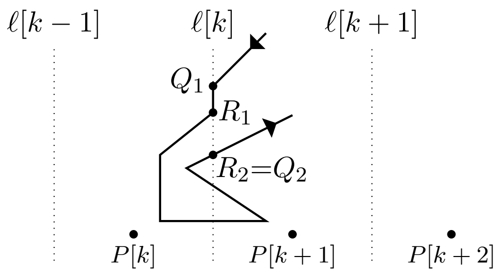

Figure 6: Excursion from the right at $\ell[k]$.

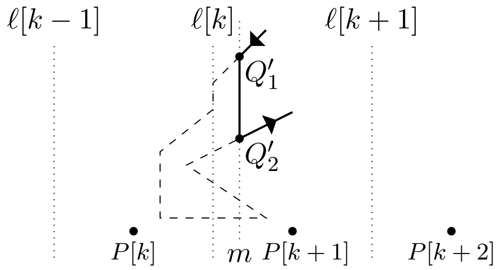

Figure 7: The excursion from Figure 6 truncated in Q'_1 and Q'_2 on line m.

sisting of line segments of non-zero length, as in the case of Q_1 and R_1, would not invalidate any of the arguments of the proof.

Excursions are the core obstacle that needs to be overcome for our proof. We can *truncate* an excursion at $\ell[k]$ by finding a suitable line m that is parallel to $\ell[k]$, some small distance away from it, between $\ell[k]$ and $P[k+1]$ for excursions from the right, and between $\ell[k]$ and $P[k]$ for excursions from the left. We further need to find points $Q'_1, Q'_2 \in C \cap m$, where $d_C(Q'_1) < d_C(Q_1)$ and $d_C(Q_2) < d_C(Q'_2)$. Because our coordinates no longer need to consist of integers, it is clear that m, Q'_1 and Q'_2 satisfying these requirements must exist.

The truncation consists in changing $sub_C(Q'_1, Q'_2)$ to become $seg(Q'_1, Q'_2)$, as illustrated by Figure 7. Note that if C is say $A[k']$, for some k', then changing the shape of C means changing the shape of $A[k'']$ for any other k'' as well; the difference between $A[k']$ and $A[k'']$ is only in the starting point $P[k']$ versus $P[k'']$.

At this time, we must allow for the possibility that for some excursions, no m, Q'_1 and Q'_2 can be found with which we can implement a truncation, if we also need to preserve the four constraints and preserve absence of self-intersections. There is a small number of possible causes. First,

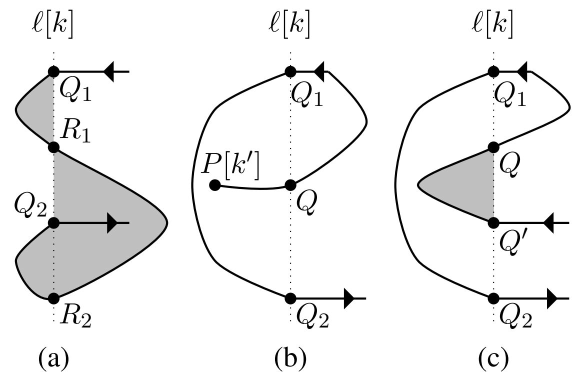

Figure 8: (a) Regions (shaded) of an excursion at $\ell[k]$; due to additional crossings in R_1 and R_2, three more excursions exist, each with a smaller area. (b) & (c) If truncation would introduce self-intersection, then either the excursion is filled, with some point $P[k']$ as in (b), or there is an excursion with smaller area, illustrated by shading in (c).

suppose that $C = A[k']$ and $B[k'+1]$ intersects with $seg(Q_1, Q_2)$. Then $B[k'+1]$ may intersect with $seg(Q'_1, Q'_2)$ for any choice of m, Q'_1 and Q'_2, and thereby no truncation is possible without violating constraint (ii). Similarly, a truncation may be blocked if $C = B[k'+1]$ and $A[k']$ intersects with $seg(Q_1, Q_2)$. Next, it could be that $C = A[k']$, while $d_{A[k']}(Q_1) > d_{A[k'+1]}(Q)$ holds for some $Q \in seg(Q_1, Q_2) \cap A[k'+1]$, or $d_{A[k'-1]}(Q) > d_{A[k']}(Q_2)$ holds for some $Q \in seg(Q_1, Q_2) \cap A[k'-1]$, either of which potentially blocks a truncation if constraint (iii) is to be preserved. Constraint (iv) has similar consequences. Furthermore, if we need to preserve absence of self-intersections, a truncation may be blocked if $d_C(Q) < d_C(Q_1)$ or $d_C(Q_2) < d_C(Q)$ for some $Q \in seg(Q_1, Q_2) \cap C$.

5 Normal form

The *regions* of an excursion of C between Q_1 and Q_2 at $\ell[k]$ are those that are enclosed by (subpaths of) $sub_C(Q_1, Q_2)$ and (subsegments of) $seg(Q_1, Q_2)$, as illustrated by Figure 8(a). The *area* of the excursion is the surface area of all regions together. We say an excursion is *filled* if any of its regions contains at least one point $P[k']$, for some integer k', otherwise it is said to be *unfilled*.

We say A and B are in *normal form* if no excursion can be truncated without violating the four constraints or introducing a self-intersection. Suppose A and B are in normal form, while one or more excursions remain. Let us first consider the unfilled excursions. Among them choose one that has the smallest area. By assumption, one of the four constraints must be violated or a new self-intersection must be introduced, if we were to truncate that excursion. We will consider all relevant cases.

Each case will assume an unfilled excursion from the right (excursions from the left are symmetric) of a path C between Q_1 and Q_2 at $\ell[k]$. We may assume that $sub_C(Q_1, Q_2) \cap \ell[k] = \{Q_1, Q_2\}$, as additional crossings of $\ell[k]$ would mean that excursions exist with smaller areas (cf. Figure 8(a)), contrary to the assumptions. Now assume truncation is blocked due to $Q \in seg(Q_1, Q_2) \cap C$ such that $d_C(Q) < d_C(Q_1)$ (the case $d_C(Q_2) < d_C(Q)$ is symmetric), as we need to preserve absence of self-intersection. Suppose Q is the only such point, so that C crosses $seg(Q_1, Q_2)$ from left to right once without ever crossing it from right to left, until Q_1 is reached. Then C starts in the area of the excursion, or in other words, the excursion is filled, contrary to the assumptions (cf. Figure 8(b)). Now suppose there are points Q' and Q where C crosses $seg(Q_1, Q_2)$ from right to left and from left to right, respectively and $d_C(Q') < d_C(Q) < d_C(Q_1)$. If there are several choices, choose Q' and Q such that $sub_C(Q', Q) \cap \ell[k] = \{Q', Q\}$. This means the excursion between Q' and Q has an area smaller than the one between Q_1 and Q_2, contrary to the assumptions (cf. Figure 8(c)).

Note that excursions with zero area, that is, those that intersect with $\ell[k]$ without crossing over to the other side, can always be truncated. We can therefore further ignore non-crossing intersections.

Now suppose a truncation would violate constraint (ii), where $C = B[k'+1]$ and $D = A[k']$ crosses $seg(Q_1, Q_2)$. Then much as above, we may distinguish two cases. In the first, D has only one crossing of $seg(Q_1, Q_2)$ in some point Q, which means the excursion is filled with the starting or ending point of D, as in Figure 9(a). In the second, D has at least two consecutive crossings, say in Q and Q', from right to left and from left to right, respectively, which means the excursion between Q and Q' has smaller area than the one between Q_1 and Q_2, illustrated by shading in Figure 9(b). Both cases contradict the assumptions.

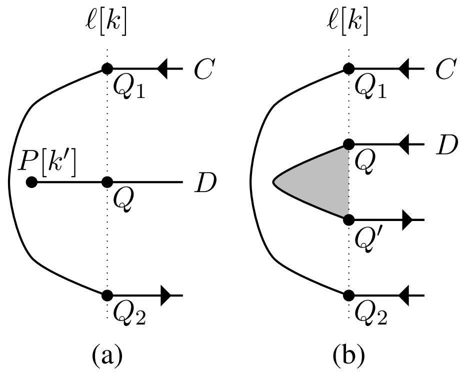

Figure 9: Truncating the excursion would introduce a violation of constraint (ii). The assumptions are contradicted in one of two ways.

For $C = A[k']$ and $D = B[k'+1]$, the reasoning is symmetric.

Next, suppose a truncation would violate constraint (iii), where $C = A[k']$ and $A[k'-1]$ crosses $seg(Q_1, Q_2)$ in Q, while $d_{A[k'-1]}(Q) > d_{A[k']}(Q_2)$. If the crossing in Q is from right to left, and there is an immediately next crossing in Q' from left to right, then we have the same situation as in Figure 9(b), involving an excursion with smaller area, contradicting the assumptions. If the crossing in Q is the only one, and it is from right to left, then we can use the fact that $sub_{A[k']}(Q_1, Q_2) \cap sub_{A[k'-1]}(Q, P[k']) = \emptyset$, as we assume the four constraints as yet hold. This means $P[k']$ must be contained in the area of the excursion, as illustrated in Figure 10(a), contradicting the assumption that the excursion is unfilled. If the crossing in Q is the only one, and it is from left to right, then we can use the fact that $sub_{A[k']}(Q_1, Q_2) \cap sub_{A[k'-1]}(Q'_2, Q) = \emptyset$, for the unique $Q'_2 \in A[k'-1] \cap \ell[k-1]$ such that $d_{A[k'-1]}(Q'_2) = d_{A[k']}(Q_2)$. This means the excursion contains Q'_2, which implies there is another unfilled excursion between points $R_1, R_2 \in A[k'] \cap \ell[k-1]$ with smaller area, as shaded in Figure 10(b), contrary to the assumptions.

Suppose a truncation would violate constraint (iii), where $C = A[k']$ and $A[k'+1]$ crosses $seg(Q_1, Q_2)$ in Q, while $d_{A[k']}(Q_1) > d_{A[k'+1]}(Q)$. The reasoning is now largely symmetric to the above, with the direction of the crossing reversed, except that the case analogous to Figure 10(b) is immediately excluded, as Q'_2 cannot be both to the left and to the right of $\ell[k]$. Constraint (iv) is symmetric to constraint (iii). All pos-

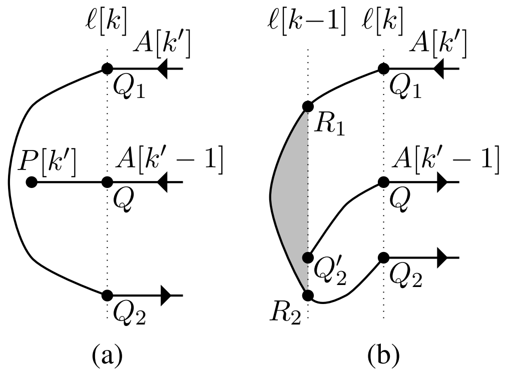

Figure 10: Truncating the excursion would introduce a violation of constraint (iii), where $d_{A[k'-1]}(Q) > d_{A[k']}(Q_2)$. The assumptions are contradicted in one of three ways, the first as in Figure 9(b), and the second and third as in (a) and (b) above.

sible cases have been shown to lead to contradictions, and therefore we conclude that there are no unfilled excursions if A and B are in normal form.

We now show that there cannot be any filled excursions either. For this, assume that $A[k']$ has a filled excursion between Q_1 and Q_2 at $\ell[k]$ from the right. This means $A[k'-1]$ has an identically shaped, filled excursion at $\ell[k-1]$ from the right, between corresponding points Q'_1 and Q'_2. Let us consider how path $A[k']$ proceeds after reaching Q_2. There are only three possibilities:

- it ends in $P[k+1]$, with $k' = k$, without further crossings of $\ell[k]$ or $\ell[k+1]$;
- it next crosses $\ell[k]$ leftward; or
- it next crosses $\ell[k+1]$ in some point Q_3.

The first of these can be excluded, in the light of $d_{A[k'-1]}(Q) \geq d_{A[k']}(Q_2)$ for each $Q \in sub_{A[k'-1]}(Q'_2, P[k'])$. Due to constraint (iii) therefore, this subpath of $A[k'-1]$ cannot intersect with the excursion of $A[k']$ to reach $P[k]$, and therefore $A[k']$ cannot reach $P[k+1]$. The second possibility is also excluded, as this would imply the existence of an unfilled excursion. For the remaining possibility, $Q_3 \in A[k'] \cap \ell[k+1]$ may be lower down than Q_2 (in the now familiar view of the points $P[0], P[1], \ldots$ being drawn from left to right along a horizontal line), or it may be higher up than Q_1. These two cases are drawn in Figures 11 and 12. The choice of Q_3 also determines a corresponding $Q'_3 \in A[k'-1] \cap \ell[k]$.

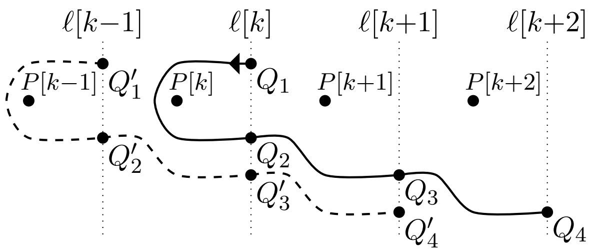

Figure 11: Continuing the (solid) path $A[k']$ after a filled excursion, restricted by the (dashed) path $A[k'-1]$, in the light of constraint (iii).

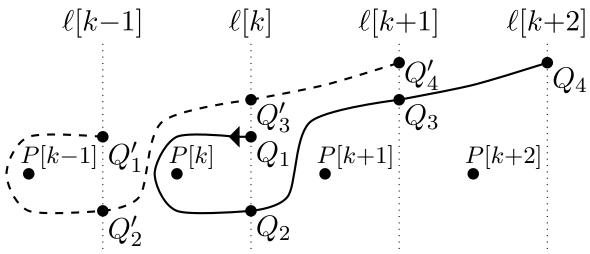

Figure 12: As in Figure 11 but Q_3 is chosen to be higher up than Q_1.

We now consider how $A[k']$ continues after Q_3 in the case of Figure 11. If it next crosses $\ell[k+1]$ leftward, this would imply the existence of an unfilled excursion. Further, $d_{A[k'-1]}(Q) \geq d_{A[k']}(Q_3)$ for each $Q \in sub_{A[k'-1]}(Q'_3, P[k'])$. Due to constraint (iii) therefore, this subpath of $A[k'-1]$ cannot intersect with $sub_{A[k']}(Q_2, Q_3)$, above which lies $P[k+1]$. Therefore, $A[k']$ must cross $\ell[k+2]$ in some Q_4, which is lower down than Q_3. This continues ad infinitum, and $A[k']$ will never reach its supposed end point $P[k'+1]$. The reasoning for Figure 12 is similar.

Filled excursions from the left are symmetric, but instead of investigating the path after Q_2, we must investigate the path *before* Q_1. The case of B is symmetric to that of A. We may now conclude no filled excursions exist.

6 The final contradiction

We have established that after A and B have been brought into normal form, there can be no remaining excursions. This means that $A[0]$ crosses $\ell[0]$ exactly once, in some point R_A, and $B[0]$ crosses $\ell[-1]$ exactly once, in some point L_B. Further, let L_A be the unique point where $A[-1]$ crosses $\ell[-1]$ and R_B the unique point where $B[1]$ crosses $\ell[0]$.

The region of the plane between $\ell[-1]$ and $\ell[0]$

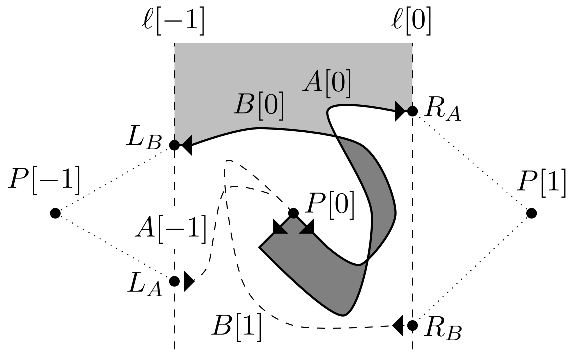

Figure 13: The region between $\ell[-1]$ and $\ell[0]$ is divided by $A[0]$ and $B[0]$ into a 'top' region (lightly shaded), a 'bottom' region (white), and areas enclosed by intersections of $A[0]$ and $B[0]$ (darkly shaded). Here $A[-1]$ and $B[1]$ are both in the 'bottom' region.

can now be partitioned into a 'top' region, a 'bottom' region, and zero or more enclosed regions. The 'top' region consists of those points that are reachable from any point between $\ell[-1]$ and $\ell[0]$ arbitrarily far above any point of $A[0]$ and $B[0]$, without intersecting with $A[0]$, $B[0]$, $\ell[-1]$ or $\ell[0]$. This is the lightly shaded region in Figure 13. The 'bottom' region is similarly defined, in terms of reachability from any point between $\ell[-1]$ and $\ell[0]$ arbitrarily far *below* any point of $A[0]$ and $B[0]$. The zero or more enclosed regions stem from possible intersections of $A[0]$ and $B[0]$; the two such enclosed regions in Figure 13 are darkly shaded. Note that the four constraints do *not* preclude intersections of $A[0]$ and $B[0]$.

However, constraint (ii) implies that, between $\ell[-1]$ and $\ell[0]$, $A[0]$ and $B[1]$ do not intersect other than in $P[0]$, and similarly, $A[-1]$ and $B[0]$ do not intersect other than in $P[0]$. Moreover, for any $Q \in sub_{A[-1]}(L_A, P[0])$ and any $Q' \in sub_{A[0]}(P[0], R_A)$ we have $d_{A[-1]}(Q) \geq d_{A[0]}(Q')$. By constraint (iii) this means $A[-1]$ and $A[0]$ do not intersect other than in $P[0]$. Similarly, $B[1]$ and $B[0]$ do not intersect other than in $P[0]$.

The angles in $P[0]$ between $A[0]$, $B[0]$, $A[-1]$ and $B[1]$ are multiples of 90°. Because of constraint (i), which excludes a 180° angle between $A[0]$ and $B[0]$, it follows that either $sub_{A[-1]}(L_A, P[0])$ and $sub_{B[1]}(R_B, P[0])$ both lie entirely in the 'top' region, or both lie entirely in the 'bottom' region. The latter case is illustrated in Figure 13. In the former case, L_A and R_B are above L_B and R_A, respectively, and in the latter case L_A and R_B are below L_B and R_A. This is impossible, as L_A and R_A should be at the same height, these being corresponding points of $A[-1]$ and $A[0]$, which have the same shape, and similarly L_B and R_B should be at the same height.

This contradiction now leads back to the very beginning of our proof, and implies that the four constraints cannot all be true, and therefore that at least one rule is always applicable to allow use of the inductive hypothesis, and therefore that G generates $\mathbf{O}_2$.

7 Conclusions and outlook

We have presented a new proof that $\mathbf{O}_2$ is generated by a MCFG. It has at least superficial elements in common with the proof by Salvati (2015). Both proofs use essentially the same MCFG, both are geometric in nature, and both involve a continuous view of paths next to a discrete view. The major difference lies in the approach to tackling the myriad ways in which the paths can wind around each other and themselves. In the case of Salvati (2015), the key concept is that of the complex exponential function, which seems to restrict the proof technique to two-dimensional geometry. In our case, the key concepts are excursions and truncation thereof, and the identification of top and bottom regions.

At this time, no proof is within reach that generalizes the result to $\mathbf{O}_3$, i.e. the language of strings over an alphabet of six symbols, in which the number of a's equals the number of $\overline{a}$'s, the number of b's equals the number of $\overline{b}$'s, and the number of c's equals the number of $\overline{c}$'s; this language is rationally equivalent to MIX-4, which is defined analogously to MIX, but with four symbols. One may expect however that a proof would use three-dimensional geometry and generalize some of the arguments from this paper. Our aim here is to make this plausible, while emphasizing that an actual proof will require a novel framework at least as involved as that presented in the previous sections.

Omitting the start rule and the axioms, an obvious candidate MCFG to generate $\mathbf{O}_3$ would among others have the three rules:

$$R(p_1q_1, p_2q_2, q_3p_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$$
$$R(p_1q_1, q_2p_2, p_3q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$$
$$R(q_1p_1, p_2q_2, p_3q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$$

as well as the six rules:

$R(p_1q_1p_2, p_3q_2, q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1q_1p_2, q_2, p_3q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1q_1, p_2q_2p_3, q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1q_1, q_2, p_2q_3p_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1, q_1p_2q_2, q_3p_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1, q_1p_2, q_2p_3q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$

Consider three strings x, y and z such that $xyz \in \mathbf{O}_3$. If we can use any of the above rules to divide these into six strings out of which we can select three, which concatenated together are a non-empty string in $\mathbf{O}_3$ shorter than xyz, then we can use the inductive hypothesis, much as in Section 2. For a proof by contradiction, therefore assume that no pair of prefixes of x and y and a suffix of z together form a non-empty string in $\mathbf{O}_3$ shorter than xyz, etc., in the light of the first three rules above, and assume that no 'short enough' prefix of x, a prefix of y and a 'short enough' suffix of x together form a non-empty string in $\mathbf{O}_3$, etc., in the light of the next six rules above.

For a geometric interpretation, consider the paths of x, y and z, leading from point $P_0 = (0, 0, 0)$ to points P_x, P_y and P_z, respectively. The concatenations of prefixes of x and y, and similarly those of x and z and those of y and z form three connecting surfaces, together forming one surface dividing the space around P_0 into an 'above' and a 'below'; cf. Figure 14. Our assumptions imply that the final parts of the paths of x, y and z from $-P_x$, $-P_y$ and $-P_z$, respectively, to P_0 should not intersect with this surface. In addition, no pair of strings from x, y and z should end on complementing symbols, i.e. a and $\overline{a}$, b and $\overline{b}$, or c and $\overline{c}$. This means that the three paths leading towards P_0 must all end in P_0 strictly 'above' or all strictly 'below' the surface.

This might lead to a contradiction, similar to that in Section 6, but only if one can ensure that none of the three paths to P_0 'sneak around' the surface. This is illustrated in Figure 15, where the path of z is 'entangled' with a copy of itself. It appears this can be achieved by adding three more rules, namely:

$R(p_1q_1p_2q_2, p_3, q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1, q_1p_2q_2p_3, q_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$
$R(p_1, q_1, q_2p_2q_3p_3) \leftarrow R(p_1, p_2, p_3)\ R(q_1, q_2, q_3)$

The physical interpretation of, say, the last rule seems to be that the path of z from $-P_z$ to P_0 can be iteratively shifted such that points other than its ending point coincide with P_0. At some stage the shifted path must intersect with the path of z from P_y to $-P_x$, before the entanglement of the two paths is broken.

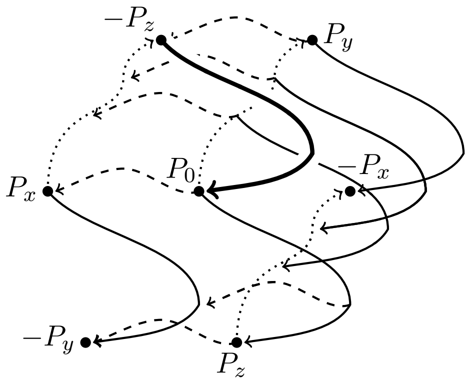

Figure 14: By taking prefixes of two strings from $\{x, y, z\}$ and concatenating them, we obtain a surface dividing the space around P_0 into 'above' and 'below'. Here the path of z from $-P_z$ to P_0 ends 'above', if our view is from above the surface.

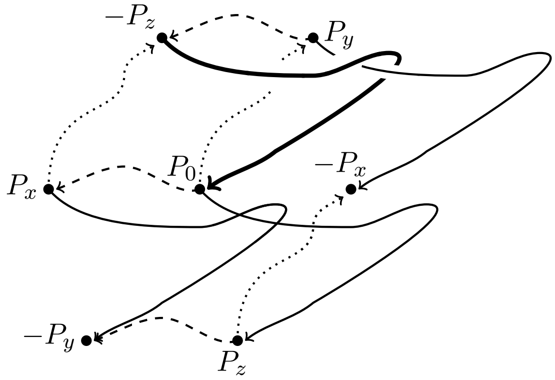

Figure 15: The path of z from $-P_z$ to P_0 is initially above the surface, but 'sneaks around' the path of z from P_y to $-P_x$, to end below.

The considerable challenges ahead involve finding a suitable definition of 'excursions' in three dimensions, and proving that these can be systematically truncated without violating appropriate constraints that preclude application of the above 12 rules.

Acknowledgements

This work came out of correspondence with Giorgio Satta. Gratefully acknowledged are also fruitful discussions with Sylvain Salvati, Vinodh Rajan, and Markus Pfeiffer. Much appreciated are anonymous referees and editors for their efforts and their courage to consider a theoretical paper for publication at this venue.

References

[Bach1981] E. Bach. 1981. Discontinuous constituents in generalized categorial grammars. In *Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistic Society*, pages 1–12.

[Boullier1999] P. Boullier. 1999. Chinese numbers, MIX, scrambling, and Range Concatenation Grammars. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 53–60, Bergen, Norway, June.

[Capelletti and Tamburini2009] M. Capelletti and F. Tamburini. 2009. Parsing with polymorphic categorial grammars. *Research in Computing Science*, 41(2009):87–98.

[Gilman2005] R.H. Gilman. 2005. Formal languages and their application to combinatorial group theory. *Contemporary Mathematics*, 378:1–36.

[Joshi et al.1991] A.K. Joshi, K. Vijay-Shanker, and D. Weir. 1991. The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S.M. Shieber, and T. Wasow, editors, *Foundational Issues in Natural Language Processing*, chapter 2, pages 31–81. MIT Press.

[Joshi1985] A.K. Joshi. 1985. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D.R. Dowty, L. Karttunen, and A.M. Zwicky, editors, *Natural language parsing: Psychological, computational, and theoretical perspectives*, pages 206–250. Cambridge University Press.

[Kallmeyer2010] Laura Kallmeyer. 2010. *Parsing Beyond Context-Free Grammars*. Springer-Verlag.

[Kanazawa and Salvati2012] M. Kanazawa and S. Salvati. 2012. MIX is not a tree-adjoining language. In *50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 666–674, Jeju Island, Korea, July.

[Pullum1983] G.K. Pullum. 1983. Context-freeness and the computer processing of human languages. In *21st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 1–6, Cambridge, Massachusetts, July.

[Salvati2015] S. Salvati. 2015. MIX is a 2-MCFL and the word problem in $\mathbb{Z}^2$ is captured by the IO and the OI hierarchies. *Journal of Computer and System Sciences*, 81:1252–1277.

[Seki et al.1991] H. Seki, T. Matsumura, M. Fujii, and T. Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science*, 88:191–229.

[Sorokin2014] A. Sorokin. 2014. Pumping lemma and Ogden lemma for displacement context-free grammars. In *Developments in Language Theory, 18th International Conference*, volume 8633 of *Lecture Notes in Computer Science*, pages 154–165, Ekaterinburg, Russia. Springer-Verlag.

[Vijay-Shanker et al.1987] K. Vijay-Shanker, D.J. Weir, and A.K. Joshi. 1987. Characterizing structural descriptions produced by various grammatical formalisms. In *25th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 104–111, Stanford, California, USA, July.