Algorithm for testing membership in a CFL

Given a CFG G and a string w, to determine whether w is generated by G:

If $w = \varepsilon$ check whether S is nullable.

If $w = a_1 a_2 \ldots . a_n$ we have to first bring G to CNF and then use an efficient $O(n^3)$ algorithm based on dynamic programming known as the CYK algorithm found independently by T. Cocke, D. Younger and T. Kasami :

Compute an upper-triangular table X_{ij} ($i \le j$) where a table-entry X_{ij} will eventually be a set of variables which generate $a_i\, a_{i+1} \ldots\ldots a_j$.

Initially for all i,j $i \le j$ $X_{ij} \leftarrow \Phi$

For the main diagonal :

for i=1 to n

for every A such that $A \to a_i$ insert A in X_{ii} .

Now the X_{ij}'s for $i < j$ will have to be computed diagonally from the main diagonal to the upper right corner ie to X_{1n} . This means that X_{ij} can be computed only after X_{ik} for $k < j$ and X_{kj} for $k > i$ are computed. X_{ij} can be computed by for k=i to j-1

if X_{ik} contains B and X_{k+1},j contains C and A -> B C is a production

insert A in X_{ij} .

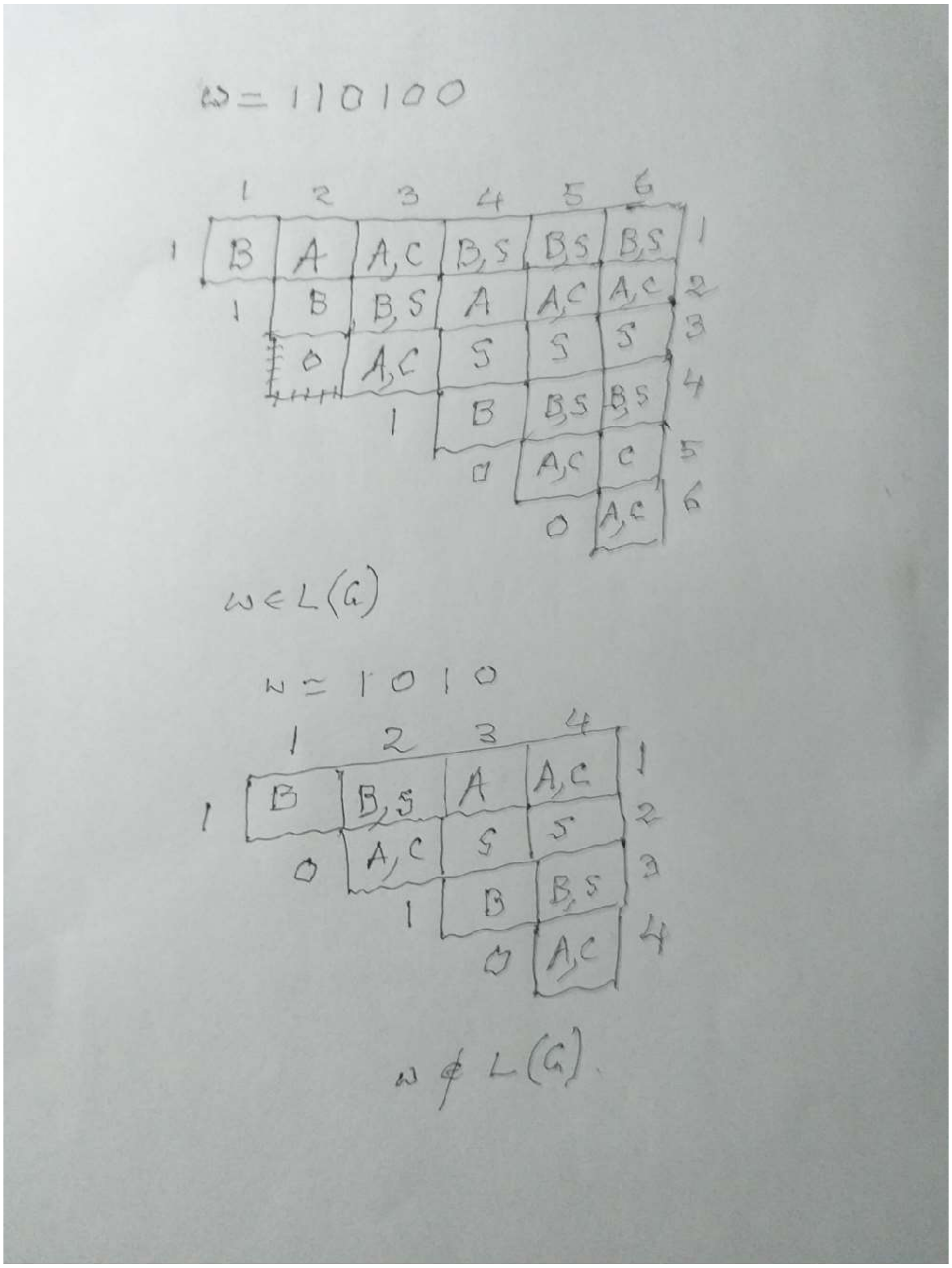
Filling up the table in this way we can determine whether G generates w by checking whether X_{1n} contains S.

Example S -> A B | B C

A -> B B | 0

B -> B A | 1

C-> A C | C A | 0



When w = 110100 we fill up the squares adjacent to the left of the main diagonal, by the string w as shown. Now since B -> 1, A -> 0 and C -> 0 we can fill up the squares X_{11}, X_{22} and X_{44} with B and the squares X_{33}, X_{55} and X_{66}

with A, C. Now we come to the next diagonal. The first entry to be computed is X_{12}. For this we need the pair of squares X_{11} and X_{22} which give a possible body of production BB. We find that this is produced by A and so we insert A in X_{12}. For X_{23} we need the pair X_{22} and X_{33} giving possible bodies of production BA and BC which are produced respectively by B and S and so we insert B and S in X_{23}. Like this we fill up the other squares of this diagonal by S; B,S; and C. This way we go on filling the squares diagonal by diagonal. For example when we come to X_{25} we need the pairs (X_{22} X_{35}), (X_{23} X_{45}), (X_{24} X_{55}) which yield possible bodies of production BS, BB, BS, SB, SS and AC. Of these only BB and AC are valid bodies of production produced by A and C respectively. So A and C get inserted in X_{25}. In this way filling up all the squares we find that X_{16} (Note that n = 6 here) contains S and hence G generates w. In the next case we take the string w = 1010. After filling up the table we find that X_{14} does not contain S and so G does not generate w.