CS3342 – Assignment 2
due Feb. 16, 2023
2-day no-penalty extension until: Feb. 18, 11:55pm

1. (30pt) Consider postfix expressions, which are arithmetic expressions in which the operator comes after the operands: `a + b` is written as `a b +`. Assume `const` operands and `$` end marker.
numbers

   (a) (10pt) Write an SLR(1) grammar, G_r, for postfix expressions, that is not LL(1). Give the FIRST and FOLLOW sets for all nonterminals. Show that G_r is not LL(1). Draw the LR parser as a graph; the states contain the LR-items, the transitions are labelled by tokens, reduce states are double circled. Include also (as jflap does) the trivial states, containing a single LR-item with the dot at the end. Build its LR parse table (as done by jflap) to prove it is SLR(1).

   (b) (5pt) Draw the parse tree (in G_r) for the string `1 2 3 4 + * - 5 / 6 7 * + $`.

   (c) (10pt) Write an LL(1) grammar, G_ℓ, for postfix expressions. Build its LL parse table (as done by jflap) to prove it is LL(1). Give also the FIRST and FOLLOW sets for all nonterminals.

   (d) (5pt) Draw the parse tree (in G_ℓ) for the string `1 2 3 4 + * - 5 / 6 7 * + $`.

2. (20pt) Scientific notation is expressing a number as $m \times 10^n$, where m, the *mantissa*, is a decimal number and n, the *exponent*, is an integer. Scientific e notation is the same thing written in a single line as m `e` n. We will assume, for simplicity, that the mantissa must have digits before and after the decimal point and that the exponent must have at least one digit. Examples: `40.5e3`, `8.0e0`, `-23.11e20`, `+0.234e-9`.

   (a) (15pt) Construct an attribute grammar that uses only one attribute, *val*; the value of *val* for the root will store the value corresponding to the number represented by the scientific e notation given by the yield of the tree.

   (b) (5pt) Draw an annotated parse tree for the string `-12.345e-10`. Show the attribute flow (arrows and values).

3. (50pt) Write a Python program, `balance.py`, which computes the minimum number of edit operations to balance a string. The *edit distance* between two string is the minimum number of operations – insertion, deletion, replacement – necessary to transform one string into another. A string consisting only of parentheses, `(` and `)` is called *balanced* if its parentheses can be properly matched: each open parenthesis, `(`, with a following closed parenthesis, `)`.

   The program, on an input string consisting only of parentheses, outputs:

   - the smallest edit distance, d, between the input string and a balanced string and
   - a balanced string at edit distance d from the input.

   For example:
   `$ python balance.py "(()())()"` outputs: `d = 0, balanced string: "(()())()"`
   `$ python balance.py ")())"` outputs: `d = 1, balanced string: "(())"`
   `$ python balance.py "))()))"` outputs: `d = 2, balanced string: "()()()"`
   `$ python balance.py ")))(("` outputs: `d = 3, balanced string: "()()"`

   Note that there can be more than one closest balanced string, and the program is allowed to output any such solution. Also, optimality proof is not required.

READ ME! Submit your answers as a *single pdf file* in OWL. Solutions should be typed; readable (by others!) hand-written solutions are also acceptable. Source code, if required, is submitted as separate files.

JFLAP: You are allowed to use JFLAP to help you solve the assignment. Make sure you understand what it does; JFLAP will not be available during in-person exams!

LaTeX: For those interested, the best (the only!) program for scientific writing is LaTeX. It is free and you can start using it in minutes: `https://tobi.oetiker.ch/lshort/lshort.pdf`

(a) (10pt) Write an SLR(1) grammar, G_r, for postfix expressions, that is not LL(1). Give the FIRST and FOLLOW sets for all nonterminals. Show that G_r is not LL(1). Draw the LR parser as a graph; the states contain the LR-items, the transitions are labelled by tokens, reduce states are double circled. Include also (as jflap does) the trivial states, containing a single LR-item with the dot at the end. Build its LR parse table (as done by jflap) to prove it is SLR(1).

①

0 P -> E $

← start state.

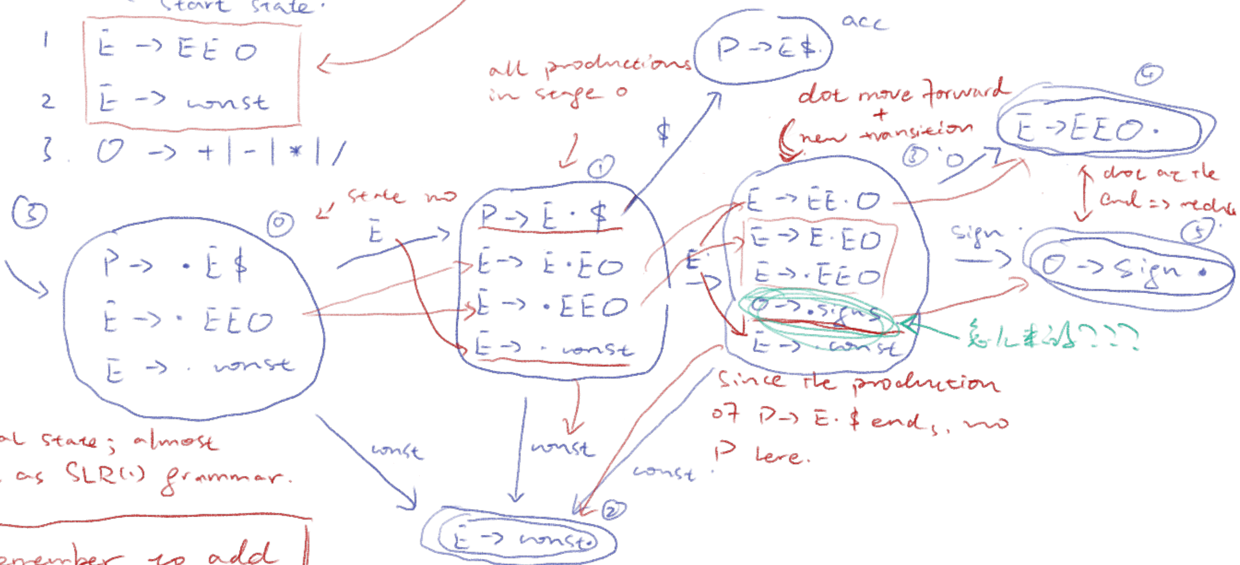
1 E -> E E O

2 E -> const

3 O -> + | - | * | /

② Since E would produce EEO and const, it is not LL(1) ← one to one production.

③



Remember to add the dot!

if no more non-terminal left, then it is a final state.

stage 0, received E, using rule 1 to move forward

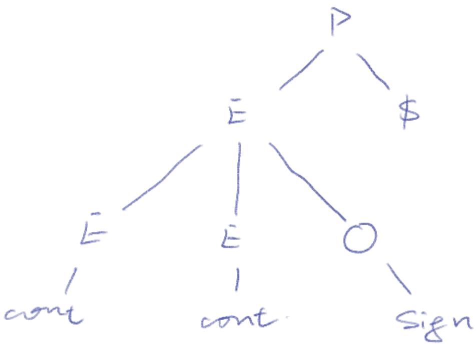
* only on final state we use reduce rule.

state \ token	sign	const	\$	E	O
0		s2		s1	
1		s2	acc.	s3.	
2	r2.	r2.	r2		
3	s5	s2.			s4
4	r1	r1	r1		
5					
6	r3.	r3	r3		
7					
8					

reduce with rule 2.

no conflict => SLR(1)

(b) (5pt) Draw the parse tree (in G_r) for the string `1 2 3 4 + * - 5 / 6 7 * + $`.



(c) (10pt) ① Write an LL(1) grammar, G_ℓ, ② for postfix expressions. Build its LL parse table (as done by jflap) to prove it is LL(1). ③ Give also the FIRST and FOLLOW sets for all nonterminals.

left-to-right input

left-to-right derivation.

=> all derivation should have const at the left.

①

0 P -> E $

1 E -> const X

2 X -> E O X

3 X -> ε

4 O -> sign

②

term \ non-term	sign	const	$
E		1	
O	4		
X	3	2	3

T(N, t) is the production rule used to expend N when lookahead is t. Blank indicates syntax error

(E, const) = 1 : E -> const X.

(O, sign) = 4 : O -> sign

(X, sign) = 3 : X -> ε.

(X, const) = 2 : X -> E O X

E -> const X

(X, $) = 3 : X -> ε.

③

	FIRST	FOLLOW
E	const	$, sign
X	const	sign, const, $
O	sign	sign, $

E -> const X

-> const E O X

-> const E sign

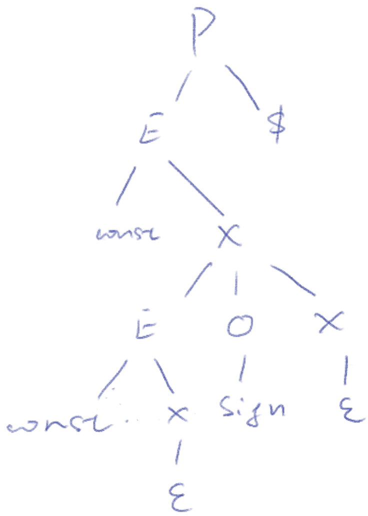
x -> EOx

-> const, sign

O -> sign.

(d) (5pt) Draw the parse tree (in G_ℓ) for the string `1 2 3 4 + * - 5 / 6 7 * + $`.

Similar with SLR(1), but using LL(1) grammar



2. (20pt) Scientific notation is expressing a number as $m \times 10^n$, where m, the *mantissa*, is a decimal number and n, the *exponent*, is an integer. Scientific e notation is the same thing written in a single line as m e n. We will assume, for simplicity, that the mantissa must have digits before and after the decimal point and that the exponent must have at least one digit. Examples: `40.5e3`, `8.0e0`, `-23.11e20`, `+0.234e-9`.

   (a) (15pt) Construct an attribute grammar that uses only one attribute, *val*; the value of *val* for the root will store the value corresponding to the number represented by the scientific e notation given by the yield of the tree.

   (b) (5pt) Draw an annotated parse tree for the string `-12.345e-10`. Show the attribute flow (arrows and values).

(a) $S \to A_1 P_1 . R \, e \, A_2 P_2$

$P \to D$

$P_1 \to P_2 D$

$R \to D R$

$R \to D$

$D \to \{0, 1, 2, 3 \cdots 9\}$.

$A \to + \mid - \mid \varepsilon$.

Step 1. give a regular

SLR(1) / LL(1)

sign   the definition of men.

$S.val \leftarrow A_1.val \times (P_1.val + R.val \times 10^{-1}) \times 10^{A_2.val \times P_2.val}$

attribute grammar => add the description of the grammar.