Prof. Ryan Cotterell

Danilo Dordevic: Assignment 3

ddordevic@student.ethz.ch, 21-954-888.

10/01/2023 - 16:07h

Question 1

a)

$$\begin{aligned} a^* &= \bigoplus_{n=0}^{\infty} a^{\otimes n} \\ &= a^{\otimes 0} \oplus \bigoplus_{n=1}^{\infty} a^{\otimes n} \\ &= a^{\otimes 0} \oplus a^{\otimes 1} \otimes \bigoplus_{n=0}^{\infty} a^{\otimes n} \\ &= \boxed{1 \oplus a \otimes a^*} \end{aligned}$$

b)

$$\begin{aligned} a^* &= \bigoplus_{n=0}^{\infty} a^{\otimes n} \\ &= \bigoplus_{n=0}^{\infty} \{a + a + \cdots + a\} \\ &= \bigoplus_{n=0}^{\infty} \{n \cdot a\} \\ &= \boxed{\log\left(\sum_{n=0}^{\infty} e^{n \cdot a}\right)} \\ &= \log\left(1 + e^a + e^{2a} + \cdots\right) \\ &= \log\left(\frac{1}{1 - e^a}\right), \quad a < 0 \end{aligned} \tag{1}$$

c) In section a) of question 1, it has been shown that a^* as defined there is in fact a general form of the Kleene star operator for a given semiring with operations $\oplus$ and

$\otimes$, since it satisfies all of the conditions necessary to be a Kleene star. Here I use the expression 1. Another note: $a \in \mathbb{R} \times \mathbb{R},\ a = (a_1, a_2),\ a_1, a_2 \in \mathbb{R}$.

$$\begin{aligned}
a^* &= \bigoplus_{n=0}^{\infty} a^{\otimes n} \\
&= a^{\otimes 0} \oplus a^{\otimes 1} \oplus a^{\otimes 2} \oplus \cdots \\
&= \mathbf{1} \oplus a \oplus (a \otimes a) \oplus (a \otimes a \otimes a) \oplus \cdots \\
&= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \oplus \begin{pmatrix} a_1^2 \\ 2a_1 a2 \end{pmatrix} \oplus \begin{pmatrix} a_1^3 \\ 3a_1^2 a2 \end{pmatrix} \oplus \begin{pmatrix} a_1^4 \\ 4a_1^3 a2 \end{pmatrix} \oplus \cdots \\
&= \boxed{\begin{pmatrix} \sum_{n=0}^{\infty} a_1^n \\ \sum_{n=1}^{\infty} n a_1^{n-1} a_2 \end{pmatrix}} \\
&= \begin{pmatrix} \frac{1}{1-a_1} \\ \frac{a_2}{(1-a_1)^2} \end{pmatrix}, \quad |a_1| < 1
\end{aligned}$$

d) Proving that $< 2^{\Sigma^*}, \cup, \otimes, \{\}, \{\varepsilon\} >$ is a semiring:

1. $< 2^{\Sigma^*}, \cup, \{\} >$ is a commutative monoid:
   - Identity element:

$$\left(\exists \{\} \in 2^{\Sigma^*}\right) \left(\forall A \in 2^{\Sigma^*}\right) : \ A \cup \{\} = \{\} \cup A = A$$

   - Associativity of $\cup$: Holds because union of sets is associative.

$$\forall A, B, C \in 2^{\Sigma^*} : \ (A \cup B) \cup C = A \cup (B \cup C)$$

   - Commutativity of $\cup$: Holds because union of sets is commutative.

$$\forall A, B \in 2^{\Sigma^*} : A \cup B = B \cup A$$

2. $< 2^{\Sigma^*}, \otimes, \{\varepsilon\} >$ is a monoid:
   - Identity element: Concatenating a string with the ε string either from left or right yields the same string.

$$\left(\exists \{\varepsilon\} \in 2^{\Sigma^*}\right) \left(\forall A \in 2^{\Sigma^*}\right) : \begin{array}{l} \{\varepsilon\} \otimes A = \{\varepsilon \circ a | a \in A\} = \{a | a \in A\} = A \\ A \otimes \{\varepsilon\} = \{a \circ \varepsilon | a \in A\} = \{a | a \in A\} = A \end{array}$$

   - Associativity of $\otimes$: The fact that the opration of string concatenation $\circ$ is associative means that $\otimes$ will also be associative.

$$\begin{aligned}
A \otimes B &= \{a \circ b | a \in A, b \in B\} \\
(A \otimes B) \otimes C &= \{(a \circ b) \circ c | a \in A, b \in B, c \in C\} \\
&= \{a \circ (b \circ c) | a \in A, b \in B, c \in C\} \\
&= A \otimes (B \otimes C)
\end{aligned}$$

3. $\otimes$ distributes over $\cup$, $\forall A, B, C \in 2^{\Sigma^*}$:

$$\begin{aligned} B \cup C = D &= \{d | d \in B \vee d \in C\} \\ A \otimes (B \cup C) &= A \otimes D \\ &= \{a \circ d | a \in A, d \in D\} \\ &= \{a \circ d | a \in A, d \in B \vee d \in C\} \\ &= \{a \circ d | a \in A, d \in B\} \cup \{a \circ d | a \in A, d \in C\} \\ &= (A \otimes B) \cup (A \otimes C) \end{aligned}$$

$$\begin{aligned} (A \cup B) \otimes C &= \{d \circ c | d \in A \vee d \in B, c \in C\} \\ &= \{d \circ c | d \in A\, c \in C\} \cup \{d \circ c | d \in B, c \in C\} \\ &= (A \otimes C) \cup (B \otimes C) \end{aligned}$$

4. $\{\}$ is the annihilator for $\otimes$: Cartesian product of a non-empty set $A \in 2^{\Sigma^*}$ with an empty set $\{\}$ is an empty set $\{\}$, i.e. it holds:

$$\forall A \in 2^{\Sigma^*} : \; A \otimes \{\} = \{\} \otimes A = \{\}$$

Kleene star of the language semiring: Applying the 1 to the language semiring, one produces the following set of expressions:

$$\begin{aligned} A^* &= \bigcup_{n=0}^{\infty} A^n \\ &= \{\varepsilon\} \cup A \cup (A \otimes A) \cup (A \otimes A \otimes A) \cup \dots \end{aligned} \tag{2}$$

To see if 2 is actually the Kleene star operation for the language semiring, it must be tested if 2 satisfies two properties:

$$\begin{aligned} A^* &= 1 \cup A \otimes A^* \\ &= \{\varepsilon\} \cup A \otimes A^* \\ &= \{\varepsilon\} \cup \{a \circ a^* | a \in A, a^* \in A^*\} \\ &= \{\varepsilon\} \cup A^* \\ &= A^* \end{aligned} \qquad \begin{aligned} A^* &= 1 \cup A^* \otimes A \\ &= \{\varepsilon\} \cup A^* \otimes A \\ &= \{\varepsilon\} \cup \{a^* \circ a | a^* \in A^*, a \in A\} \\ &= \{\varepsilon\} \cup A^* \\ &= A^* \end{aligned}$$

Having proven that 2 satisfies both properties, it can be concluded that A^* as defined in 2 is the Kleene star operator for the language semiring. Moreover, for the language semiring it holds that the Kleene star of an element of 2^{Σ^*} is actually the Kleene closure of that set.

Question 2

a) Tropical semiring is 0-closed:

$$\forall a \in \mathbb{R}_{\geq 0} : \; \min(a, 0) = 0, \text{ since } a \geq 0$$

Arctic semiring is 0-closed:

$$\forall a \in \mathbb{R}_{\leq 0} : \; \max(a, 0) = 0, \text{ since } a \leq 0$$

b) This can be proven by induction. Note: $\Pi(i,j,n)$ denotes a set of all paths of length n, starting at node i and ending at node j.

- Base case $n = 1$: $M_{i,k}$ contains the sum of path weights of all paths of length 1 from node i to node k, by definition of the matrix M. If the path exists, i.e. if nodes i and k are adjacent, the matrix will contain the weight of the corresponding edge. If the nodes are not connected, the weight is 0.
- Induction hypothesis for $n \geq 1$: $M^n_{i,k}$ contains the sum of all paths of length n that start at node i and end at node k.
- Induction step for $n + 1$:

$$\begin{aligned} M^{n+1}_{i,k} &= \bigoplus_{j=1}^{N} M_{i,j} \otimes M^n_{j,k} \\ &= \bigoplus_{j=1}^{N} w_{i,j} \otimes \left(\bigoplus_{p \in \Pi(j,k,n)} w^{(p)}_{j,k} \right) \\ &= \bigoplus_{j=1}^{N} \bigoplus_{p \in \Pi(j,k,n)} w_{i,j} \otimes w^{(p)}_{j,k} \end{aligned}$$

  $M^n_{j,k}$ contains the sum of all weights of paths of length n starting at node j and ending at node k. That can be represented by the sum of the weights $w^{(p)}_{j,k}$ of all paths p that connect nodes j and k and have length n. The product $w_{i,j} \otimes w^{(p)}_{j,k}$ represents the weight of the path of length n+1, starting at node i, going to node j and following the path p, which ends at node k. The two sums basically represent the sums of all such paths, starting at node j. This is exactly what $M^{n+1}_{i,k}$ contains by definition.

This concludes the inductive proof of the the statement that $M^n_{i,k}$ contains the semiring-sum of all paths from the node i to node k in graph G of length exactly n.

c) Analysing the first row of equation 3, it can be noticed that the sum is over all paths that start at node i and end at node j, regardless of their length. This sum can be written in a different way, such that the sum goes over all paths starting at length 0. This means that the sum is now over the values $M^n_{i,j}$ for $n \geq 0$.

$$\begin{aligned} Z(i,j) &= \bigoplus_{\pi \in \Pi(i,j)} w(\pi) \\ &= \bigoplus_{n=0}^{\infty} M^n_{i,j} \end{aligned} \tag{3}$$

Terms that correspond to paths that have length larger than N must necessarily contain cycles. The spanning tree of a graph with N nodes contains $N - 1$ branches, and it by definition contains no cycles, so adding a new edge would create a cycle. This means that some edge weights will appear as a factor in $w(\pi)$ more than once. Those weights, corresponding to the non-cycling paths, can be pulled out in front of the semiring-sum, thanks to the distribution property of the semiring, leaving inside a sum of a weight term corresponding to the cycle and 1, which would result in 1 annihilating the cycle weight term. This way it was shown that any path of length larger or equal to N does not contribute to the value $Z(i,j)$, and only terms corresponding to non-cyclic paths (meaning paths of length $\leq N-1$) contribute to it. An illustrative example of this is shown on figure 1:

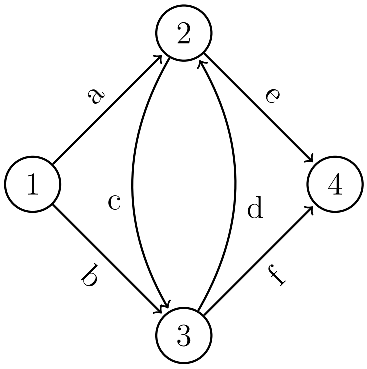

Figure 1: Illustrative example

Consider two paths from node 1 to node 4 in the graph shown on figure 1 - path π_1: 1→2→4, and path π_2: 1→2→3→2→4. The weights of these path will be included in the sum 3. The weights of paths π_1 and π_2 are:

$$\begin{aligned}w(\pi_1) &= a \otimes e \\ w(\pi_2) &= a \otimes c \otimes d \otimes e\end{aligned}$$

When summed, these path weights will yield the following:

$$\begin{aligned}w(\pi_1) + w(\pi_2) &= a \otimes e \oplus a \otimes c \otimes d \otimes e \\ &= (a \otimes e) \otimes (\underbrace{1 \oplus c \otimes d}_{1}) \\ &= (a \otimes e) \otimes 1 \\ &= a \otimes e \\ &= w(\pi_1)\end{aligned}$$

This way it was proved that paths of length $\geq N$ do not influence the sum in 3. Another way to look at this is that in the tropical or arctic semiring, paths of length $\geq N$ will never have a smaller or larger weight than the corresponding paths of length $\leq N-1$, so the min or max operation will always disregard them in the sum 3.

d) In the previous point it was shown that summands for $n \geq N$ in the equation in the second row of 3 do not influence the total sum, so 3 can be reformulated in matrix form as following:

$$\begin{aligned}Z &= \bigoplus_{n=0}^{\infty} M^n \\ &= \bigoplus_{n=0}^{N-1} M^n = M^*\end{aligned}$$

The expression for Z shown here is the Kleene star, by definition of the Kleene star, and general properties shown in 1.

e) The simple algorithm consist of computing M^n, $\forall n \in \{1, 2, \ldots, N-1\}$, starting from $M^0 = 1$. Furthermore, the only variables needed is the current variable that contains the latest matrix product M^n, and another one that contains the matrix M. The steps would look like the following:

$$\begin{array}{ll} M^0, & M^* \leftarrow M^0 \\ M^1 = M^0 \otimes M, & M^* \leftarrow M^* \oplus M^1 \\ M^2 = M^1 \otimes M, & M^* \leftarrow M^* \oplus M^1 \\ \vdots & \vdots \\ M^n = M^{n-1} \otimes M, & M^* \leftarrow M^* \oplus M^1 \end{array}$$

Each of N steps contains one $N \times N$ matrix multiplication, which, if we assume a naive implementation, is performed in $O(N^3)$, and one matrix addition, which is done in $O(N^2)$. Since this is performed N times, the final computational complexity of computing M^n is $\boxed{O(N^4)}$.

f) Showing that a 0-closed semiring is idempotent:

$$\begin{aligned} \forall a \in A: \; a &= a \otimes 1 \\ &= a \otimes (1 \oplus 1) \\ &= (a \otimes 1) \oplus (a \otimes 1) \\ &= a \oplus a \end{aligned}$$

Transition from first to second row is done by exploiting the fact that in a 0-closed semiring it holds for all items $a \in A$ that $1 \oplus a = a$. Since 1 is also in A, this property hold for it as well. Step from 2 to 3 exploit the distributive property of the semiring, and the steps from 3 to 4 exploit the neutral element 1 for the $\otimes$ operation.

g) The proof starts with a different formulation of the main semiring-sum:

$$\bigoplus_{n=0}^{K} M^n \tag{4}$$

Notice that 4 can be written in the following way:

$$\bigoplus_{n=0}^{K} M^n = (I \oplus M) \otimes (I \oplus \bigoplus_{n=0}^{K-1} M^n) \tag{5}$$

The proof for the equation 5 exploits the property of idempotent semirings that the semiring-addition of the element from the semiring with itself is just that element, and is following:

$$
\begin{aligned}
(I \oplus M) \otimes (I \oplus \bigoplus_{n=0}^{K-1} M^n) &= I \oplus \bigoplus_{n=0}^{K-1} M^n \oplus M \oplus \bigoplus_{n=0}^{K-1} M^{n+1} \\
&= I \oplus \bigoplus_{n=0}^{K-1} M^n \oplus M \oplus \bigoplus_{n=1}^{K} M^n \\
&= I \oplus I \oplus \bigoplus_{n=1}^{K-1} M^n \oplus M \oplus \bigoplus_{n=1}^{K-1} M^n \oplus M^K \\
&= I \oplus \bigoplus_{n=1}^{K-1} M^n \oplus M^K \\
&= \bigoplus_{n=0}^{K} M^n
\end{aligned}
$$

Having shown this, we can exploit the recursive structure of equation 5 in the following way:

$$
\begin{aligned}
\bigoplus_{n=0}^{K} M^n &= (I \oplus M) \otimes (I \oplus \bigoplus_{n=0}^{K-1} M^n) \\
&= (I \oplus M) \otimes (I \oplus M) \otimes (I \oplus \bigoplus_{n=0}^{K-2} M^n) \\
&= (I \oplus M)^2 \otimes (I \oplus \bigoplus_{n=0}^{K-2} M^n) \\
&\vdots \\
&= (I \oplus M)^K \otimes (I \oplus \underbrace{\bigoplus_{n=0}^{0} M^n}_{M^0 = I}) \\
&= (I \oplus M)^K \otimes (I \oplus I) \\
&= (I \oplus M)^K \otimes I \\
&= \boxed{(I \oplus M)^K}
\end{aligned}
$$

h) Rewriting the expression for M^n, we produce the following array of expressions:

$$
\begin{aligned}
M^n &= \bigotimes_{k=0}^{\lfloor \log_2 n \rfloor} M^{\alpha_k 2^k} && (6) \\
&= M^{\alpha_0} \otimes M^{2\alpha_1} \otimes M^{4\alpha_2} \otimes \cdots \otimes M^{2^{\lfloor \log_2 n \rfloor} \alpha_{\lfloor \log_2 n \rfloor}} && (7) \\
&= M^{\alpha_0 + 2\alpha_1 + 4\alpha_2 + \cdots + 2^{\lfloor \log_2 n \rfloor} \alpha_{\lfloor \log_2 n \rfloor}} && (8) \\
&= M^{\sum_{k=0}^{\lfloor \log_2 n \rfloor} \alpha_k 2^k} && (9)
\end{aligned}
$$

From the first and last row, it can be concluded that all that is necessary for the equalities to hold is for the exponents to match, i.e. it need to hold:

$$n = \sum_{k=0}^{\lfloor \log_2 n \rfloor} \alpha_k 2^k$$

The sum on the right is the expression for converting a binary number $\alpha_{\lfloor \log_2 n \rfloor}\alpha_{\lfloor \log_2 n \rfloor - 1} \dots \alpha_1\alpha_0$ into its decimal representation n, or equivalently, the way to find a decimal number n's binary representation. To represent a decimal number n in a binary format, $\lfloor \log_2 m \rfloor$ binary digits are needed.

Since for all numbers $n \in \mathbb{Z}_{\geq 0}$ there exists a binary representation $\alpha_{\lfloor \log_2 n \rfloor}\alpha_{\lfloor \log_2 n \rfloor - 1} \dots \alpha_1\alpha_0$, the equation 6 always holds, and the proof is concluded.

Now, to find a faster algorithm than the one in Q2 e), the expression for M^* needs to be rewritten, exploiting the idempotency property, Q2 f) and Q2 g):

$$\begin{aligned}
M* &= \sum_{n=0}^{N-1} M^n && (10)\\
&= (I+M)^{N-1} && (11)\\
&= \bigotimes_{k=0}^{\lfloor \log_2 (N-1) \rfloor} (I+M)^{\alpha_k 2^k} && (12)\\
&= \underbrace{(I+M)^{\alpha_0} \otimes (I+M)^{2\alpha_1} \otimes \dots \otimes (I+M)^{2^{\lfloor \log_2 (N-1) \rfloor}\alpha_{\lfloor \log_2 (N-1) \rfloor}}}_{\lfloor \log_2 (N-1) \rfloor \text{ terms}} && (13)
\end{aligned}$$

The efficient algorithm would consist of computing all of the terms from above, with the assumption that all α_i are equal to 1. In other words, all of the powers of 2 from above of matrix $(I + M)$ are computed. First, the algorithm starts with $I + M$, then in the next step computes $(I + M)^2$ by multiplying the previous matrix $I + M$ by itself, which takes $O(N^3)$. For each term we do the same - to get the current matrix, we multiply the previous one by itself. Since there are $\lfloor \log_2(N-1) \rfloor$ terms, and computing each one takes $O(N^3)$, the total complexity of computing all terms is $O(N^3 \log_2 N)$. However, this is not the end. Now, it needs to be checked which of the $\lfloor \log_2(N-1) \rfloor$ terms need to be multiplied together, as in 13. Only the terms with corresponding alphas that are equal to 1 are multiplied. The worse case scenario is when $\alpha_0 = \alpha_1 = \dots = \alpha_{\lfloor \log_2 (N-1) \rfloor} = 1$. In that case, to compute M^*, we need additional $\lfloor \log_2(N-1) \rfloor - 1$ matrix multiplications, which brings the total complexity of computing M^* to $\boxed{O(N^3(\log_2 N)^2)}$.

i) Take the SVD of matrix A:

$$A = USV^T$$

Matrices U and V are orthogonal, meaning that $UU^T = U^TU = I$ and $VV^T = V^TV = I$, so they do not change the norm of a vector when they are multiplied by it, i.e. it holds:

$$\|Ux\|_2^2 = x^TU^TUx = x^Tx = \|x\|_2^2$$

The squared 2-norm of the product of matrix A and a vector x is:

$$\begin{aligned}
\|Ax\|_2^2 &= \|USV^Tx\|_2^2 && (14)\\
&= \|S\underbrace{V^Tx}_{y}\|_2^2 && (15)\\
&= \|Sy\|_2^2 && (16)\\
&= \|\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\|_2^2 && (17)\\
&= \|\begin{pmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_n y_n \end{pmatrix}\|_2^2 && (18)\\
&= \sum_{i=1}^{n} \sigma_i^2 y_i^2 && (19)
\end{aligned}$$

The 2-norm of the product of matrix A and a vector x is:

$$\|Ax\|_2 = \sqrt{\sum_{i=1}^{n} \sigma_i^2 y_i^2}, \text{ where } y = Ax \text{ and } \|y\|_2 = \|x\|_2 \tag{20}$$

It holds that the maximal σ value can be extracted from 20 by taking the supremum of the expression 20 with respect to the vector y, under the constrain that it is a vector of unit length. This works because all of the values under the sum are non-negative and because taking the supremum will cause all of the 'mass' of vector y, which is 1, to be placed on the coordinate j that corresponds to the largest value σ_j.

$$\begin{aligned}
\sigma_{max} &= \sup_{\|y\|_2=1} \sqrt{\sum_{i=1}^{n} \sigma_i^2 y_i^2} && (21)\\
&= \sup_{\|y\|_2\neq 0} \frac{\sqrt{\sum_{i=1}^{n} \sigma_i^2 y_i^2}}{\|y\|_2} && (22)\\
&= \sup_{\|x\|_2\neq 0} \frac{\|Ax\|_2}{\|y\|_2} && (23)
\end{aligned}$$

Here the third equality holds because the numerator is exactly the norm of the product of matrix A and x, and the denominator is has exactly the same 2-norm as vector x, since y is just an orthogonal transformation of vector x, which keeps its norm. This was shown in 20.

j)

$$\|A^* - \sum_{n=0}^{K} A^n\|_2 = \sup_{x\neq 0} \frac{\|A^*x - \sum_{n=0}^{K} A^n x\|_2}{\|x\|_2} \tag{24}$$

$$= \sup_{x\neq 0} \frac{\|\sum_{n=0}^{\infty} A^n x - \sum_{n=0}^{K} A^n x\|_2}{\|x\|_2} \tag{25}$$

$$= \sup_{x\neq 0} \frac{\|\sum_{n=K+1}^{\infty} A^n x\|_2}{\|x\|_2} \tag{26}$$

$$\leq \sup_{x\neq 0} \frac{\sum_{n=K+1}^{\infty} \|A^n x\|_2}{\|x\|_2} \tag{27}$$

$$= \sum_{n=K+1}^{\infty} \sup_{x\neq 0} \frac{\|A^n x\|_2}{\|x\|_2} \tag{28}$$

$$= \sum_{n=K+1}^{\infty} \sigma_{max}(A^n) \tag{29}$$

$$= \sum_{n=K+1}^{\infty} (\sigma_{max}(A))^n \tag{30}$$

$$= \boxed{\frac{(\sigma_{max}(A))^{K+1}}{1-\sigma_{max}(A)}}\ ,\ |\sigma_{max}(A)| < 1 \tag{31}$$

$$\xrightarrow{K\to\infty} 0 \tag{32}$$

The aboslute value of the maximal singular value of matrix A needs to be strictly less than 1 in order for the infinite sum 30 to converge. The same condition is needed for the approximation error to go to 0, when K goes to ∞.

k) The bound for the approximation error depends on the number of summands K exponentially:

$$\|A^* - \sum_{n=0}^{K} A^n\|_2 = \frac{(\sigma_{max}(A))^{K+1}}{1-\sigma_{max}(A)} = \boxed{O((\sigma_{max}(A))^K)} \tag{33}$$

Provided that the maximal singular value of matrix A is less than 1, the approximation decreases exponentially, so the conclusion is that the using a truncation as a method of approximation of asteration is very good, since the error decreases significantly even with a small K.

Question 3

https://colab.research.google.com/drive/1jPitqq7svo6iV9-ank5GaEDcD6ty7PWA?usp=sharing