*Prof. Ryan Cotterell*

# Simon Wachter: Assignment 3

siwachte@ethz.ch, 19-920-198

28/12/2022 - 13:10h

## Question 1:

a) We show that under Definition 1, the following holds:

$$a^* \overset{!}{=} 1 \oplus a \otimes a^* \tag{1}$$
$$= 1 \oplus a \otimes \bigotimes_{n=0}^{\infty} a^{\otimes n} \tag{2}$$
$$= 1 \oplus \bigotimes_{n=0}^{\infty} a^{\otimes n+1} \tag{3}$$
$$= 1 \oplus \bigotimes_{n=1}^{\infty} a^{\otimes n} \tag{4}$$
$$= \bigotimes_{n=0}^{\infty} a^{\otimes n} \tag{5}$$
$$= a^* \tag{6}$$

b) First we show that $\log a \oplus a = \log(2) + a$:

$$a \oplus \log a = \log(e^a + e^a) \tag{7}$$
$$= \log(2e^a) \tag{8}$$
$$= \log(2) + a \tag{9}$$

Then we calculate the Kleene start:

$$\bigoplus_{\log\ n=0}^{\infty} a^{\oplus n} = a^{\oplus 0} \oplus_{\log} \left( \bigoplus_{\log\ n=1}^{\infty} a^{\oplus n} \right) \tag{10}$$
$$= 0 \oplus_{\log} \left( \bigoplus_{\log\ n=1}^{\infty} a^{\oplus n} \right) \tag{11}$$
$$= 0 \oplus_{\log} a \oplus_{\log} 2a \oplus_{\log} 3a \oplus_{\log} \cdots \tag{12}$$
$$= \log\left(e^0 + e^a\right) \oplus_{\log} 2a \oplus_{\log} 3a \oplus_{\log} \cdots \tag{13}$$
$$= \log\left(e^{\log(e^0+e^a)} + e^{2a}\right) \oplus_{\log} 3a \oplus_{\log} \cdots \tag{14}$$
$$= \log\left(e^0 + e^a + e^{2a}\right) \oplus_{\log} 3a \oplus_{\log} \cdots \tag{15}$$
$$= \log\left(\sum_{n=0}^{\infty} e^{a^{\oplus n}}\right) \tag{16}$$

We have two cases here, either $a > 0$ or $a \leq 0$. In the first case, the sum diverges and we get $\infty$. In the second case, the sum converges:

$$\sum_{n=0}^{\infty} e^{a^{\oplus n}} = \frac{1}{1 - e^a} \qquad \text{limit geometric series} \tag{17}$$

Therefore, we get:

$$\begin{aligned}
\log\left(\sum_{n=0}^{\infty} e^{a^{\oplus n}}\right) &= \log\left(\frac{1}{1 - e^a}\right) && (18)\\
&= \log(1) - \log(1 - e^a) && (19)\\
&= \log(1 - e^a) && (20)
\end{aligned}$$

c) First we derive a closed form solution for $(x, y)^{\oplus n}$:

$$\begin{aligned}
\langle x, y\rangle^{\oplus 1} &= \langle x, y\rangle && (21)\\
\langle x, y\rangle^{\oplus 2} &= \langle x^2, 2xy\rangle && (22)\\
\langle x, y\rangle^{\oplus 3} &= \langle x^3, 3x^2y\rangle && (23)\\
\langle x, y\rangle^{\oplus 4} &= \langle x^4, 4x^3y\rangle && (24)\\
\langle x, y\rangle^{\oplus n} &= \langle x^n, nx^{n-1}y\rangle && (25)
\end{aligned}$$

Then we derive a closed form for $a^*$:

$$\begin{aligned}
\bigoplus_{n=0}^{\infty} \langle x, y\rangle^{\oplus n} &= \bigoplus_{n=0}^{\infty} \langle x^n, nx^{n-1}y\rangle && \text{eq. (25)} \quad (26)\\
&= \left\langle \sum_{n=0}^{\infty} x^n, \sum_{n=0}^{\infty} nx^{n-1}y \right\rangle && (27)
\end{aligned}$$

Both parts only converge for $|x| < 1$. The left parts is a geometric series and has limit $\frac{1}{1-x}$. And the right parts, which is a power series:

$$\begin{aligned}
\sum_{n=0}^{\infty} nx^{n-1}y &= y\sum_{n=0}^{\infty} nx^{n-1} && (28)\\
&= y\left(0 + \sum_{n=1}^{\infty} nx^{n-1}\right) && (29)\\
&= y\sum_{n=0}^{\infty}(n+1)x^n && (30)\\
&= y\frac{1}{(x-1)^2} && \text{limit power series} \quad (31)\\
&= \frac{y}{(x-1)^2} && (32)
\end{aligned}$$

Hence our closed form with inserted limits is given by:

$$\left\langle \sum_{n=0}^{\infty} x^n, \sum_{n=0}^{\infty} nx^{n-1}y \right\rangle = \left\langle \frac{1}{1-x}, \frac{y}{(x-1)^2} \right\rangle \qquad |x| \leq 1 \tag{33}$$

If $|x| > 1$ both parts diverge.

d)

$$\mathcal{W}_{\text{lang}} = \left\langle 2^{\Sigma^*}, \bigcup, \otimes, \{\}, \{\epsilon\} \right\rangle \tag{34}$$

We first show that $\mathcal{W}_{\text{lang}}$ is a semiring:

- $(2^{\Sigma^*}, \oplus, \mathbf{0})$ must be a commutative monoid with identity element $\mathbf{0}$:

$$\begin{aligned} (x \oplus y) \oplus z &= (x \cup y) \cup z && (35)\\ &= \{x, y\} \cup z && (36)\\ &= \{x, y, z\} && (37)\\ &= x \oplus \{y, z\} && (38)\\ &= x \oplus (y \oplus z) && (39) \end{aligned}$$

$$\begin{aligned} \mathbf{0} \oplus x &= \{\} \oplus x && (40)\\ &= \{\} \cup x && (41)\\ &= \oplus && (42) \end{aligned}$$

$$\begin{aligned} x \oplus y &= x \cup y && (43)\\ &= \{x, y\} && (44)\\ &= y \cup x && (45)\\ &= y + x && (46) \end{aligned}$$

- $(2^{\Sigma^*}, \otimes, \mathbf{1})$ must be a monoid with identity element $\mathbf{1}$:

$$\begin{aligned} (x \otimes y) \otimes z &= xy \otimes z && (47)\\ &= xyz && (48)\\ &= x \otimes yz && (49)\\ &= x \otimes (y \otimes z) && (50) \end{aligned}$$

$$\begin{aligned} \mathbf{1} \otimes x &= \{\epsilon\} \otimes x && (51)\\ &= x && (52)\\ &= x \otimes \{\epsilon\} && (53)\\ &= x \otimes \mathbf{1} && (54) \end{aligned}$$

- Multiplication left and right distributes over addition:

$$\begin{aligned} x \otimes (y \oplus z) &= x \otimes \{y, z\} && (55)\\ &= \{xy, xz\} && (56)\\ &= \{xy\} \cup \{xz\} && (57)\\ &= (x \otimes y) \oplus (x \otimes z) && (58) \end{aligned}$$

$$\begin{aligned} (x \oplus y) \otimes z &= \{x, y\} \otimes z && (59)\\ &= \{xz, yz\} && (60)\\ &= \{xz\} \cup \{yz\} && (61)\\ &= (x \otimes z) \oplus (y \otimes z) && (62) \end{aligned}$$

- Multiplication by $\mathbf{0}$ annihilates $\mathbb{R} \times \mathbb{R}$:

$$\begin{aligned}
\mathbf{0} \otimes x &= \{a \circ b \,|\, a \in A, b \in \{\}\} && (63)\\
&= \{\} && \text{by definition of } \circ\text{, because no } b \text{ exists} \quad (64)\\
&= x \otimes \mathbf{0} && (65)\\
& && (66)
\end{aligned}$$

The Kleene start for $\mathcal{W}_{\text{lang}}$ given by:

$$A^{\otimes n} = \{\bigotimes_{i=0}^{n} a_i \mid a_i \in A\} \tag{67}$$

$$\begin{aligned}
A^* &= \bigoplus_{n=0}^{\infty} A^{\otimes n} && (68)\\
&= \bigoplus_{n=0}^{\infty} \left\{ \bigotimes_{i=0}^{n} a_i \mid a_i \in A \right\} && (69)\\
&= \left\{ \bigotimes_{i=0}^{n} a_i \mid a_i \in A, n \in \mathbb{Z} \right\} && (70)
\end{aligned}$$

## Question 2:

a) Tropical semiring is 0-closed:

$$\begin{aligned}
a \oplus \mathbf{0} &= \min(a, \mathbf{0}) && (71)\\
&= \min(a, 0) && (72)\\
&= 0 && \text{because } a \in \mathbb{R}_{\geq 0} \quad (73)
\end{aligned}$$

Arctic semiring is 0-closed:

$$\begin{aligned}
a \oplus \mathbf{0} &= \max(a, \mathbf{0}) && (74)\\
&= \max(a, 0) && (75)\\
&= 0 && \text{because } a \in \mathbb{R}_{\leq 0} \quad (76)
\end{aligned}$$

b) Proof by induction:
Base case $i = 1$:

$$\begin{aligned}
M^1 &= M && (77)\\
M_{ij} &= w_{ij} && \text{by def of } M \quad (78)
\end{aligned}$$

$w_{ij}$ is exactly the semiring-sum over all paths from $i$ to $j$ of length 1. This holds because there is only one path of length 1 from $i$ to $j$.
Induction hypothesis: $M^i_{ij}$ is the semiring-sum over all paths from $i$ to $j$ of length $i$.
Induction step $i \to i+1$:

$$\begin{aligned}
M^{i+1} &= M^i \otimes M && (79)\\
M^{i+1}_{kj} &= \sum_{l=0}^{n} M^i_{kl} \otimes M_{lj} && \text{def. matrix mult.} \quad (80)
\end{aligned}$$

In eq. (80) we sum over the product of all possible paths of length $i$ from $k$ to another node $l$ and all possible paths of length 1 from nodes $l$ to $j$. This sum is exactly the semiring-sum over all possible paths of length $i+1$ from $k$ to $j$.

c) If we have a graph $G$ with $N$ vertices, then a path with length $l \geq N$ must visit at least one vertex $v$ twice. From this follows, that we can just remove this cycle at vertex $v$ and reduce the path to a path of length at most $N-1$. If there are multiple cycles, we can remove all of them until we arrive at a path of length at most $N-1$. Let us now assume that the longest path in $G$ has length $l \geq N-1$. The weight of this path is the following:

$$v_0 \overset{w_0}{\rightarrow} v_1 \rightarrow \dots \rightarrow v_k \rightarrow \dots \rightarrow v_k \rightarrow \dots \overset{v_{l-2}}{\rightarrow} v_{l-1} \overset{w_{l-1}}{\rightarrow} v_l \tag{81}$$
$$w_0 \otimes \dots \otimes w_{k-1} \otimes w_k \otimes \dots \otimes w'_k \otimes w_{k+1} \otimes \dots \otimes w_{l-1} \tag{82}$$

Notice the cycle in the middle, which we know exists given our reasoning before. We can remove this cycle and arrive at the following path with new weight:

$$v_0 \overset{w_0}{\rightarrow} v_1 \rightarrow \dots \rightarrow v_k \rightarrow \dots \overset{v_{l-2}}{\rightarrow} v_{l-1} \overset{w_{l-1}}{\rightarrow} v_l \tag{83}$$
$$w_0 \otimes \dots \otimes w_{k-1} \otimes w_{k+1} \otimes \dots \otimes w_{l-1} \tag{84}$$

We first define weights for subpaths:

$$s_0 = w_0 \otimes \dots \otimes w_{k-1} \qquad \text{path to } k \tag{85}$$
$$s_1 = w_k \otimes \dots \otimes w'_k \qquad \text{cycle at } k\text{-th vertex} \tag{86}$$
$$s_2 = w_{k+1} \otimes \dots \otimes w_{l-1} \qquad \text{path from } k \text{ to } l \tag{87}$$
$$\tag{88}$$

We now rewright out path weights and add them over our semiring:

$$(s_0 \otimes s_1 \otimes s_2) \oplus (s_0 \otimes s_2) \tag{89}$$
$$(s_0) \otimes ((s_1 \otimes s_2) \oplus s_2) \tag{90}$$
$$(s_0) \otimes (s_2 \otimes (s_1 \oplus 1)) \tag{91}$$
$$(s_0) \otimes (s_2 \otimes 1) \qquad \text{def. 0-closed} \tag{92}$$
$$s_0 \otimes s_2 \qquad \text{def. } \mathbf{1} \tag{93}$$

We see that eq. (93) is exactly the weight of the path without the cycle. Since this holds for every path of length $l \geq N$ we can conclude that the longest path in $G$ has length at most $N-1$.

d) Per definition of the Kleene start, we have:

$$M^* = \bigoplus_{i=0}^{\infty} M^{\otimes i} \tag{94}$$

In b) we have shown that $M^{\otimes i}$ is the semiring-sum over all paths of length $i$ and in c) we have shown that the shortest path depends only on paths of length $l \leq N-1$. We also showed that under the $\bigoplus$ operation, only paths of length $l \leq N-1$ are considered. Therefore, we know that $M^*$ depends only on:

$$M^* = \bigoplus_{i=0}^{N-1} M^{\otimes i} \tag{95}$$

e) We define a simple algorithm:

**Algorithm 1:** Matrix multiplication for Kleene star

```
M;
M' ← M;
M* ← 0;
for i = 0 to N − 1 do
    for j = 0 to len(M) do
        for k = 0 to len(M) do
            for l = 0 to len(M) do
                M'_{j,k} ← M'_{j,k} ⊕ (M_{j,l} ⊗ M_{l,k})
            M*_{j,k} ← M*_{j,k} ⊕ M'_{j,k}
    M ← M';
```

The inner for loops calculate the matrix multiplication, $M^{\otimes n}$, and the outer loop iterates $N-1$ times to calculate the Kleene star. Since each loop has $\mathcal{O}(N)$ iterations, the algorithm has a runtime of $\mathcal{O}(N^4)$.

f)

$$\begin{aligned} a \oplus a &= a \otimes (\mathbf{1} \oplus \mathbf{1}) && && (96)\\ &= a \otimes \mathbf{1} && \text{def. 0-closed} && (97)\\ &= a && \text{def. } \mathbf{1} && (98) \end{aligned}$$

g) We show given equality with an induction proof: Base case $n = 0$:

$$\begin{aligned} \bigoplus_{n=0}^{0} M^n &= M^0 && (99)\\ &= (I \oplus M)^0 && (100) \end{aligned}$$

Where we assumed that:

$$I \oplus M = M \tag{101}$$

Our induction hypothesis is:

$$\bigoplus_{n=0}^{n} M^n = (I \oplus M)^n \tag{102}$$

Now for the inductive step we have $n \rightarrow n+1$:

$$\bigoplus_{i=0}^{n+1} M^i = \bigoplus_{i=0}^{n} M^i \oplus M^{n+1} \tag{103}$$

$$= M^0 \oplus \bigoplus_{i=0}^{n} M^i \oplus M^{n+1} \tag{104}$$

$$= M^0 \oplus \bigoplus_{i=0}^{n} \left(M^i \oplus M^i\right) \oplus M^{n+1} \qquad \text{(def. Idempotent)} \tag{105}$$

$$= \bigoplus_{i=0}^{n} M^i \oplus \bigoplus_{i=1}^{n+1} M^i \tag{106}$$

$$= \bigoplus_{i=0}^{n} M^i \otimes (\mathbf{I} \oplus M) \qquad \text{(def. distributive)} \tag{107}$$

$$= (\mathbf{I} \oplus M)^n \otimes (\mathbf{I} \oplus M) \qquad \text{(def. I.H.)} \tag{108}$$

$$= (\mathbf{I} \oplus M)^{n+1} \tag{109}$$

h) With the log factor, we immediately think about binary representation. We use the product of power rule to rewrite our left side of the equation:

$$\bigotimes_{k=0}^{\lfloor \log_2 n \rfloor} M^{\alpha_k 2^k} = M^{\sum_{k=0}^{\lfloor \log_2 n \rfloor} \alpha_k 2^k} \tag{110}$$

We now analyze the exponent of $M$ more closely. If we choose $\alpha_k$ to represent the k-th bit in the binary representation of $n$, we can see that $\sum_{k=0}^{\lfloor \log_2 n \rfloor} \alpha_k 2^k = n$. Therefore, we can rewrite the equation as:

$$\bigotimes_{k=0}^{\lfloor \log_2 n \rfloor} M^{\alpha_k 2^k} = M^{\sum_{k=0}^{\lfloor \log_2 n \rfloor} \alpha_k 2^k} \tag{111}$$

$$= M^n \tag{112}$$

With this insight, we can rewrite the algorithm to calculate $M^*$:

**Algorithm 2:** Matrix multiplication for Kleene star

$M$;
$M' \leftarrow M$;
$M^* \leftarrow \mathbf{0}$;
**for** $i = 0$ *to* $\log_2(N-1)$ **do**
  **if** $a_i == 0$ **then**
    **for** $j = 0$ *to len(M)* **do**
      **for** $k = 0$ *to len(M)* **do**
        **for** $l = 0$ *to len(M)* **do**
          $M'_{j,k} \leftarrow M'_{j,k} \oplus (M_{j,l} \otimes M_{l,k})$
        $M^*_{j,k} \leftarrow M^*_{j,k} \oplus M'_{j,k}$
    $M \leftarrow M'$;

As we have just remove some iterations from the outer most loop, our runtime changes to $\mathcal{O}(n^3 \log n)$

i) We derive the SVD of $A$:

$$A = U\Sigma V^T \tag{113}$$
$$||A||_2 = ||U\Sigma V^T||_2 = ||\Sigma||_2 \tag{114}$$

Further we can rewrite the operator norm by w.l.o.g choosing $x$ such that $||x||_2 = 1$:

$$\begin{aligned} ||A||_2 &= \sup_{x \neq 0} \frac{||Ax||_2}{||x||_2} && (115) \\ &= \sup_{x\neq 0, ||x||_2 = 1} ||Ax||_2 && (116) \\ & && (117) \end{aligned}$$

Combining these two insight we get:

$$\begin{aligned} ||A||_2 &= \sup_{x\neq 0, ||x||_2 = 1} ||\Sigma x||_2 && (118) \\ &= \sigma_{\max}(A) && \text{(min-max theorem)} \quad (119) \end{aligned}$$

j)

$$\begin{aligned} ||A^* - \sum_{n=0}^{K} A^n||_2 &= ||\sum_{n=0}^{\infty} A^n - \sum_{n=0}^{K} A^n||_2 && (120) \\ &= ||\sum_{n=K+1}^{\infty} A^n||_2 && (121) \\ &= \sigma_{\max}\left(\sum_{n=K+1}^{\infty} A^n\right) && \text{(ex. i))} \quad (122) \\ &\leq \sum_{n=K+1}^{\infty} \sigma_{\max}\left(A^n\right) && \text{(singular value inequalities)} \quad (123) \\ &\leq \sum_{n=K+1}^{\infty} \sigma_{\max}\left(A\right)^n && \text{(singular value inequalities)} \quad (124) \\ &= \frac{\sigma_{\max}(A)^{K+1}}{1 - \sigma_{\max}(A)} && \text{(geom. series if } \sigma_{\max}(A) < 1\text{))} \quad (125) \\ & && (126) \end{aligned}$$

eq. (125) shows the closed form solution if $\sigma_{\max}(A) < 1$. If $\sigma_{\max}(A) \geq 1$, the closed form solution is not defined as the geometric series diverges.
The condition on $\sigma_{\max}(A)$ for the truncation error to converge to $A^*$ therefore is $\sigma_{\max}(A) < 1$:

$$\lim_{K\to\infty} \frac{\sigma_{\max}(A)^{K+1}}{1 - \sigma_{\max}(A)} = 0 \tag{127}$$

k) If $\sigma_{\max}(A) \geq 1$ the truncation error is unbounded and the truncation is not a good approximation to asteration.
In the case where $\sigma_{\max}(A) < 1$ the truncation error in $\mathcal{O}$ is given by:

$$\mathcal{O}\left(\frac{\sigma_{\max}(A)^{K+1}}{1-\sigma_{\max}(A)}\right) = \mathcal{O}\left(\sigma_{\max}(A)^{K}\right) \tag{128}$$

Hence the error decays exponentially, which is then generally deemed to be acceptable for an error and so a truncation is a good approximation to asteration.