A. Proof of Theorem 3

Our strategy is to form a Jordan decomposition of A and show that the desired bounds hold for each Jordan block. To this end, we first prove the following lemmas.

Lemma 8. *If J is a Jordan block with nonzero eigenvalue, then for any $\epsilon > 0$ there is a complex matrix D such that $J + D$ is diagonalizable in $\mathbb{C}$ and*

$$\frac{\|(J+D)^n - J^n\|}{\|J^n\|} \leq n\epsilon.$$

Proof. The powers of J look like

$$J^n = \begin{bmatrix} \binom{n}{0}\lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \cdots \\ & \binom{n}{0}\lambda^n & \binom{n}{1}\lambda^{n-1} & \cdots \\ & & \binom{n}{0}\lambda^n & \cdots \\ & & & \ddots \end{bmatrix}.$$

More concisely,

$$[J^n]_{jk} = \begin{cases} \binom{n}{k-j}\lambda^{n-k+j} & \text{if } 0 \leq k-j \leq n \\ 0 & \text{otherwise.} \end{cases}$$

We choose D to perturb the diagonal elements of J towards zero; that is, let D be a diagonal matrix whose elements are in $[-\epsilon\lambda, 0)$ and are all different. This shrinks the diagonal elements by a factor no smaller than $(1-\epsilon)$. So the powers of $(J+D)$ are, for $0 \leq k-j \leq n$:

$$\begin{aligned} [(J+D)^n]_{jk} &= c_{jk}\,[J^n]_{jk} \\ c_{jk} &\geq (1-\epsilon)^{n-k+j}. \end{aligned}$$

Simplifying the bound on c_{jk} (Kozma, 2019):

$$c_{jk} \geq 1 - (n-k+j)\epsilon \geq 1 - n\epsilon. \tag{4}$$

The elements of J^n, for $0 \leq k-j \leq n$, are perturbed by:

$$\begin{aligned} [(J+D)^n - J^n]_{jk} &= (c_{jk}-1)[J^n]_{jk} \\ \left|[(J+D)^n - J^n]_{jk}\right| &\leq n\epsilon \left|[J^n]_{jk}\right|. \end{aligned}$$

Since $\|\cdot\|$ is monotonic,

$$\begin{aligned} \|(J+D)^n - J^n\| &\leq n\epsilon\,\|J^n\| \\ \frac{\|(J+D)^n - J^n\|}{\|J^n\|} &\leq n\epsilon. \end{aligned}$$

□

Lemma 9. *If J is a Jordan block with zero eigenvalue, then for any $\epsilon > 0, r > 0$, there is a complex matrix D such that $J + D$ is diagonalizable in $\mathbb{C}$ and*

$$\|(J+D)^n - J^n\| \leq r^n\epsilon.$$

Proof. Since the diagonal elements of J are all zero, we can't perturb them toward zero as in Lemma 8; instead, let

$$\delta = \min\left\{\frac{r}{2}, \left(\frac{r}{2}\right)^d \frac{\epsilon}{d}\right\}$$

and let D be a diagonal matrix whose elements are in $(0, \delta]$ and are all different. Then the elements of $((J+D)^n - J^n)$ are, for $0 \leq k-j < \min\{n, d\}$:

$$\begin{aligned} [(J+D)^n - J^n]_{jk} &\leq \binom{n}{k-j}\delta^{n-k+j} \\ &< 2^n\delta^{n-k+j} \\ &\leq 2^n\delta^{\min\{0,n-d\}+1}, \end{aligned}$$

and by monotonicity,

$$\|(J+D)^n - J^n\| \leq 2^n\delta^{\min\{0,n-d\}+1}d.$$

To simplify this bound, we consider two cases. If $n \leq d$,

$$\begin{aligned} \|(J+D)^n - J^n\| &= 2^n\delta d \\ &\leq 2^n\left(\frac{r}{2}\right)^d \frac{\epsilon}{d}d \\ &= 2^{n-d}r^d\epsilon \\ &\leq r^n\epsilon. \end{aligned}$$

If $n > d$,

$$\begin{aligned} \|(J+D)^n - J^n\| &= 2^n\delta^{n-d+1}d \\ &\leq 2^n\delta^{n-d}\left(\frac{r}{2}\right)^d \frac{\epsilon}{d}d \\ &\leq 2^n\left(\frac{r}{2}\right)^{n-d}\left(\frac{r}{2}\right)^d \frac{\epsilon}{d}d \\ &= r^n\epsilon. \end{aligned}$$

□

Now we can combine the above two lemmas to obtain the desired bounds for a general matrix.

Proof of Theorem 3. Form the Jordan decomposition $A = PJP^{-1}$, where

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix}$$

and each J_j is a Jordan block. Let $\kappa(P) = \|P\|\|P^{-1}\|$ be the Frobenius condition number of P.

If A is nilpotent, use Lemma 9 on each block J_j to find a D_j so that $\|(J_j + D_j)^n - J_j^n\| \leq \frac{r^n\epsilon}{\kappa(P)p}$. Combine the D_j

into a single matrix D, so that $\|(J+D)^n - J^n\| \leq \frac{r^n \epsilon}{\kappa(P)}$. Let $E = PDP^{-1}$, and then

$$\begin{aligned}\|(A+E)^n - A^n\| &= \|P((J+D)^n - J^n)P^{-1}\| \\ &\leq \kappa(P)\|(J+D)^n - J^n\| \\ &\leq \kappa(P)\frac{r^n \epsilon}{\kappa(P)} \\ &= r^n \epsilon.\end{aligned}$$

If A is not nilpotent, then for each Jordan block J_j:

- If J_j has nonzero eigenvalue, use Lemma 8 to find a D_j such that $\|(J_j + D_j)^n - J_j^n\| \leq \frac{n\epsilon}{\kappa(P)^2}\frac{\|J^n\|}{2p}$.
- If J_j has zero eigenvalue, use Lemma 9 to find a D_j such that $\|(J_j + D_j)^n - J_j^n\| \leq \frac{n\epsilon}{\kappa(P)^2}\frac{\rho(J)^n}{2p}$.

Combine the D_j into a single matrix D. Then the total absolute error of all the blocks with nonzero eigenvalue is at most $\frac{n\epsilon}{\kappa(P)^2}\frac{\|J^n\|}{2}$. And since $\rho(J)^n \leq \|J^n\|$, the total absolute error of all the blocks with zero eigenvalue is also at most $\frac{n\epsilon}{\kappa(P)^2}\frac{\|J^n\|}{2}$. So the combined total is

$$\|(J+D)^n - J^n\| \leq \frac{n\epsilon}{\kappa(P)^2}\|J^n\|.$$

Finally, let $E = PDP^{-1}$, and

$$\begin{aligned}\|(A+E)^n - A^n\| &= \|P((J+D)^n - J^n)P^{-1}\| \\ &\leq \kappa(P)\|((J+D)^n - J^n)\| \\ &\leq \frac{n\epsilon}{\kappa(P)}\|J^n\| \\ &\leq \frac{n\epsilon}{\kappa(P)}\|P^{-1}A^nP\| \\ &\leq n\epsilon\|A^n\| \\ \frac{\|(A+E)^n - A^n\|}{\|A^n\|} &\leq n\epsilon.\end{aligned}$$

□

B. Proof of Proposition 5

First, consider the $\oplus$ operation. Let $\mu_1(a)$ (for all a) be the transition matrices of M_1. For any $\epsilon > 0$, let $E_1(a)$ be the perturbations of the $\mu_1(a)$ such that $\|E_1(a)\| \leq \epsilon/2$ and the $\mu_1(a) + E_1(a)$ (for all a) are simultaneously diagonalizable. Similarly for M_2. Then the matrices $(\mu_1(a) + E_1(a)) \oplus (\mu_2(a) + E_2(a))$ (for all a) are simultaneously diagonalizable, and

$$\begin{aligned}&\|(\mu_1(a) + E_1(a)) \oplus (\mu_2(a) + E_2(a)) - \mu_1(a) \oplus \mu_2(a)\| \\ &\quad= \|E_1(a) \oplus E_2(a)\| \\ &\quad\leq \|E_1(a)\| + \|E_2(a)\| \\ &\quad\leq \epsilon.\end{aligned}$$

Next, we consider the $\sqcup$ operation. Let d_1 and d_2 be the number of states in M_1 and M_2, respectively. Let $E_1(a)$ be the perturbations of the $\mu_1(a)$ such that $\|E_1(a)\| \leq \epsilon/(2d_2)$ and the $\mu_1(a) + E_1(a)$ are simultaneously diagonalizable by some matrix P_1. Similarly for M_2.

Then the matrices $(\mu_1(a) + E_1(a)) \sqcup (\mu_2(a) + E_2(a))$ (for all a) are simultaneously diagonalizable by $P_1 \otimes P_2$. To see why, let $A_1 = \mu_1(a) + E_1(a)$ and $A_2 = \mu_2(a) + E_2(a)$ and observe that

$$\begin{aligned}&(P_1 \otimes P_2)(A_1 \sqcup A_2)(P_1 \otimes P_2)^{-1} \\ &\quad= (P_1 \otimes P_2)(A_1 \otimes I + I \otimes A_2)(P_1^{-1} \otimes P_2^{-1}) \\ &\quad= P_1 A_1 P_1^{-1} \otimes I + I \otimes P_2 A_2 P_2^{-1} \\ &\quad= P_1 A_1 P_1^{-1} \sqcup P_2 A_2 P_2^{-1},\end{aligned}$$

which is diagonal.

To show that $(\mu_1(a) + E_1(a)) \sqcup (\mu_2(a) + E_2(a))$ is close to $(\mu_1(a) \sqcup \mu_2(a)$, observe that

$$\begin{aligned}&(\mu_1(a) + E_1(a)) \sqcup (\mu_2(a) + E_2(a)) \\ &\quad= (\mu_1(a) + E_1(a)) \otimes I + I \otimes (\mu_2(a) + E_2(a)) \\ &\quad= \mu_1(a) \otimes I + E_1(a) \otimes I + I \otimes \mu_2(a) + I \otimes E_2(a) \\ &\quad= (\mu_1(a) \sqcup \mu_2(a)) + (E_1(a) \sqcup E_2(a)).\end{aligned}$$

Therefore,

$$\begin{aligned}&\|(\mu_1(a) + E_1(a)) \sqcup (\mu_2(a) + E_2(a)) - \mu_1(a) \sqcup \mu_2(a)\| \\ &\quad= \|E_1(a) \sqcup E_2(a)\| \\ &\quad= \|E_1(a) \otimes I + I \otimes E_2(a)\| \\ &\quad\leq \|E_1(a) \otimes I\| + \|I \otimes E_2(a)\| \\ &\quad\leq \|E_1(a)\|d_2 + d_1\|E_2(a)\| \\ &\quad\leq \epsilon.\end{aligned}$$

C. Proof of Proposition 7

Because any set of commuting matrices can be simultaneously triangularized by a change of basis, assume without loss of generality that M's transition matrices are upper triangular, that is, there are no transitions from state q to state r where $q > r$.

Let $M = (Q, \Sigma, \lambda, \mu, \rho)$, and arbitrarily number the symbols of Σ as $a_1, \ldots, a_m$. Note that M assigns the same weight to multiset w as it does to the sorted symbols of w. That is, we can compute the weight of w by summing over sequences of states $q_0, \ldots, q_m$ such that q_0 is an initial state, q_m is a final state, and M can get from state q_{i-1} to q_i while reading a_i^k, where k is the number of occurrences of a_i in w.

For all $a \in \Sigma, q, r \in Q$, define $M_{q,a,r}$ to be the automaton that assigns to a^k the same weight that M would going from state q to state r while reading a^k. That is,

$$\begin{aligned}
M_{q,a,r} &= (\lambda_{q,a,r}, \mu_{q,a,r}, \rho_{q,a,r}) \\
[\lambda_{q,a,r}]_q &= 1 \\
\mu_{q,a,r}(a) &= \mu(a) \\
[\rho_{q,a,r}]_r &= 1
\end{aligned}$$

and all other weights are zero.

Then we can build a multiset automaton equivalent to M by combining the $M_{q,a,r}$ using the union and shuffle operations:

$$M' = \bigoplus_{\substack{q_0,\ldots,q_m \in Q \\ q_0 \le \cdots \le q_m}} \lambda_{q_0} M_{q_0,a_1,q_1} \sqcup \cdots \sqcup M_{q_{m-1},a_m,q_m} \rho_{q_m}$$

(where multiplying an automaton by a scalar means scaling its initial or final weight vector by that scalar). The $M_{q,a,r}$ are unary, so by Proposition 5, the transition matrices of M' are ASD. Since $M_{q,a,r}$ has $r - q + 1$ states, the number of states in M' is

$$|Q'| = \sum_{q_0 \le \cdots \le q_m} \prod_{i=1}^{m} (q_i - q_{i-1} + 1)$$

which we can find a closed-form expression for using generating functions. If $p(z)$ is a polynomial, let $[z^i](p(z))$ stand for "the coefficient of z^i in p." Then

$$\begin{aligned}
|Q'| &= \left[z^{d-1}\right] \left(\sum_{i=0}^{\infty} z^i\right) \left(\sum_{i=0}^{\infty} (i+1) z^i\right)^m \left(\sum_{i=0}^{\infty} z^i\right) \\
&= \left[z^{d-1}\right] \left(\frac{1}{1-z}\right) \left(\frac{1}{1-z}\right)^{2m} \left(\frac{1}{1-z}\right) \\
&= \left[z^{d-1}\right] \left(\frac{1}{1-z}\right)^{2m+2} \\
&= \binom{2m+d}{d-1}.
\end{aligned}$$