B APPENDIX 2 (CONTENTS SEPARATED FROM SUBMISSION)

B.1 Example

Consider the mixture of $L = 2$ chains on $n = 4$ shown in Figure 10. States $\{1, 2\}$ and $\{3, 4\}$ are connected in the first chain whereas the second chain is fully connected. For ease of presentation, we

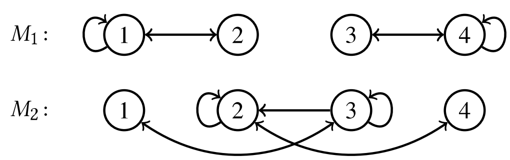

Figure 10: A Mixture of Markov Chains. Transition probabilities are uniform among all outgoing neighbors.

assume that starting probabilities are uniform (i.e. $s_j^\ell = 1/8$ for all $\ell \in \{1, 2\}$ and $j \in \{1, 2, 3, 4\}$), even though that violates the last condition of Theorem 1. We collect the likelihood of observing 3-trails $i \to j \to k$ in matrices O_j, whereas $O_j(i, k) = p(i, j, k)$:

$$O_1 = \frac{1}{8}\begin{pmatrix} 1/4 & 1/4 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad O_2 = \frac{1}{8}\begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 \\ 0 & 1/6 & 0 & 1/6 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \quad \ldots$$

These can be broken down into products of the matrices containing the 2 chains' transition probabilities

$$P_1 = \frac{1}{8}\begin{pmatrix} 1/2 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \end{pmatrix}, \quad P_2 = \frac{1}{8}\begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/3 & 0 \end{pmatrix}, \quad \ldots$$
$$Q_1 = \frac{1}{8}\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad Q_2 = \frac{1}{8}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}, \quad \ldots$$

as $O_j = \overline{P_j} \cdot 8 \cdot Q_j$ for all states $j \in \{1, 2, 3, 4\}$.

We want to convince ourselves that this mixture fulfills the first two necessary conditions for Theorem 1. We first need to ensure companion-connectivity of the chain. As such, we form the bipartite graphs G^1 and G^2 as shown in Figure 11. Clearly, 2 and 4 are the companion states of 1 and 3 (and vice versa). Next, we need to

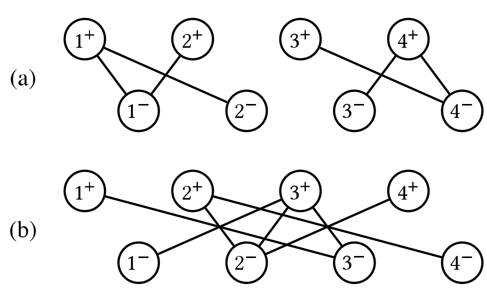

Figure 11: The bipartite graphs G^1 (a) and G^2 (b).

analyze the shuffle matrix $\mathcal{A}$ and its co-kernel. If we order the columns of $\mathcal{A}$ lexicographically, we have

$$\mathcal{A} = \begin{array}{c} \\ 1^+ \\ 2^+ \\ 3^+ \\ 4^+ \\ 1^- \\ 2^- \\ 3^- \\ 4^- \end{array} \begin{array}{cccccccccccccccc} (1,1) & (1,2) & (1,3) & (1,4) & (2,1) & (2,2) & (2,3) & (2,4) & (3,1) & (3,2) & (3,3) & (3,4) & (4,1) & (4,2) & (4,3) & (4,4) \\ P_1(1) & & & & P_1(2) & & & & P_1(3) & & & & P_1(4) & & & \\ & P_2(1) & & & & P_2(2) & & & & P_2(3) & & & & P_2(4) & & \\ & & P_3(1) & & & & P_3(2) & & & & P_3(3) & & & & P_3(4) & \\ & & & P_4(1) & & & & P_4(2) & & & & P_4(3) & & & & P_4(4) \\ \multicolumn{4}{c}{-Q_1} & & & & & & & & & & & & \\ & & & & \multicolumn{4}{c}{-Q_2} & & & & & & & & \\ & & & & & & & & \multicolumn{4}{c}{-Q_3} & & & & \\ & & & & & & & & & & & & \multicolumn{4}{c}{-Q_4} \end{array}$$

Plugging in P_j and Q_i from above, we can clearly see that each column (i, j) just contains the transition probabilities from state i to j, in each of the two chains. We can verify numerically that the co-kernel of $\mathcal{A}$ is spanned by indicator vectors for each of the three connected components

$$\xi_1 = \xi^1_{\{1,2\}}, \quad \xi_2 = \xi^1_{\{3,4\}}, \quad \xi_3 = \xi^2_{\{1,2,3,4\}}.$$

Specifically, these indicator vectors and the indicator matrices Ξ_1, Ξ_2, Ξ_3, Ξ_4 are defined as

$$\begin{array}{c} \\ \begin{pmatrix} \xi_1^\top \\ \xi_2^\top \\ \xi_3^\top \end{pmatrix} = \end{array} \begin{array}{cc:cc:cc:cc:cc:cc:cc:cc} \multicolumn{2}{c}{1^+} & \multicolumn{2}{c}{2^+} & \multicolumn{2}{c}{3^+} & \multicolumn{2}{c}{4^+} & \multicolumn{2}{c}{1^-} & \multicolumn{2}{c}{2^-} & \multicolumn{2}{c}{3^-} & \multicolumn{2}{c}{4^-} \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}$$
$$= \begin{pmatrix} \Xi_1 & \Xi_2 & \Xi_3 & \Xi_4 & \Xi_1 & \Xi_2 & \Xi_3 & \Xi_4 \end{pmatrix}$$

Our guess of the co-kernel of $\mathcal{A}'$ is correct up to a full-rank matrix $R \in \mathbb{R}^{r \times r}$. The rows of R and columns of R^{-1} correspond to the connected components $\{1, 2\}$, $\{3, 4\}$, and $\{1, 2, 3, 4\}$, in this order. Let ρ_1, ρ_2, ρ_3 be the rows of R and $\rho_1^{-1}, \rho_2^{-1}, \rho_3^{-1}$ be the columns of R^{-1}. The mixture has the two companionship equivalence classes $\{1, 2\}$ and $\{3, 4\}$. Note that indeed, $\Xi_1 = \Xi_2$ and $\Xi_3 = \Xi_4$. Assume we pick the representatives 1 and 3 with companions 2 and 4, respectively. Note that

$$R_{\mathrm{inv}(1)} = \begin{pmatrix} \rho_1^{-1} & \rho_3^{-1} \end{pmatrix} \qquad \text{and} \qquad R_{\mathrm{inv}(3)} = \begin{pmatrix} \rho_2^{-1} & \rho_3^{-1} \end{pmatrix}$$

contain only the columns corresponding to connected components that contain the states 1 or 3, respectively. Eventually, we obtain $\tilde{R}_1$ and $\tilde{R}_3$ through an eigendecomposition and correcting a scaling. Assume that the permutations in $\tilde{R}_1$ and $\tilde{R}_3$ are such that

$$\tilde{R}_1^+ = \begin{pmatrix} \rho_3^{-1} & \rho_1^{-1} \end{pmatrix} \qquad \text{and} \qquad \tilde{R}_3^+ = \begin{pmatrix} \rho_2^{-1} & \rho_3^{-1} \end{pmatrix}.$$

We combine columns from both matrices in arbitrary order and take the inverse, which yields the matrix

$$\Pi R = \begin{pmatrix} \rho_2^{-1} & \rho_3^{-1} & \rho_1^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} \rho_2 \\ \rho_3 \\ \rho_1 \end{pmatrix}.$$

Let us denote the rows of Y_1 as $Y_1(1), Y_1(2)$ and the rows of Y_3 as $Y_3(1), Y_3(2)$. We can then write the result of the recovery as

$$\Pi R Y_1' = \begin{pmatrix} 0 \\ y_1(2) \\ y_1(1) \end{pmatrix} \qquad \text{and} \qquad \Pi R Y_3' = \begin{pmatrix} y_3(1) \\ y_3(2) \\ 0 \end{pmatrix}.$$

This tells us that state 1 is part of the connected components corresponding to rows 2 and 3, and state 2 is part of the connected components corresponding to rows 1 and 2. We therefore have to label rows 1 and 3 with $\ell = 1$ and row 2 with $\ell = 2$. Compressing the matrices along this labeling then yields

$$\Pi' Y_1 = \begin{pmatrix} y_1(1) \\ y_1(2) \end{pmatrix}, \qquad \Pi' Y_2 = \begin{pmatrix} y_2(1) \\ y_2(2) \end{pmatrix}, \qquad \ldots$$

from which we can recover $\mathcal{M}$ (up to the permutation Π').

B.2 Omitted Proofs

THEOREM 2. *We have*

$$\sigma_L(P_j) \cdot \sigma_L(M_j^+) \leq \sigma_L(O_j) \leq \sqrt{L} \min \left\{ \sigma_L(P_j), \sigma_L(M_j^+) \right\}.$$

PROOF. First, note that $Q_j = S_j M_j^+$ and thus

$$O_j = \overline{P_j} S^{-1} Q_j = \overline{P_j} M_j^+.$$

We begin by showing the first inequality. We can write the L-th largest singular value as

$$\begin{aligned} \sigma_L(\overline{P_j} M_j^+) &= \max_{\dim(U)=L} \min_{u \in U} \frac{\|\overline{P_j} M_j^+ u\|_2}{\|u\|_2} \\ &= \max_{\dim(U)=L} \min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \cdot \frac{\|\overline{P_j} M_j^+ u\|_2}{\|M_j^+ u\|_2} \end{aligned}$$

where U is any subspace of R^n. We bound the above by splitting the product into

$$\begin{aligned} \sigma_L(\overline{P_j} M_j^+) &\geq \max_{\dim(U)=L} \left(\min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \right) \cdot \left(\min_{u \in U} \frac{\|\overline{P_j} M_j^+ u\|_2}{\|M_j^+ u\|_2} \right) \\ &\geq \max_{\dim(U)=L} \left(\min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \right) \cdot \left(\min_{w \in \mathbb{R}^L} \frac{\|\overline{P_j} w\|_2}{\|w\|_2} \right) \end{aligned}$$

since $M_j^+ v \in \mathbb{R}^L$. Note that the second term is just the smallest (or L-th) singular value of P_j. Hence,

$$\begin{aligned} \sigma_L(\overline{P_j} M_j^+) &\geq \max_{\dim(U)=L} \left(\min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \right) \cdot \sigma_L(P_j) \\ &= \sigma_L(P_j) \cdot \max_{\dim(U)=L} \min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \\ &= \sigma_L(P_j) \cdot \sigma_L(M_j^+) \end{aligned}$$

where U is any subspace of R^n.

For the second inequality, we upper bound

$$\|\overline{P_j} M_j^+ u\|_2 \leq \|P_j\|_2 \cdot \|M_j^+ u\|_2.$$

Since $P_j \in [0,1]^{L \times n}$, the operator norm $\|P_j\|_2$ is maximized when all entries of P_j are 1. As such, $\|P_j\|_2 \leq \sqrt{L}$. We can now bound the L-th largest singular value by

$$\begin{aligned} \sigma_L(\overline{P_j} M_j^+) &= \max_{\dim(U)=L} \min_{u \in U} \frac{\|\overline{P_j} M_j^+ u\|_2}{\|u\|_2} \\ &\leq \|P_j\|_2 \cdot \max_{\dim(U)=L} \min_{u \in U} \frac{\|M_j^+ u\|_2}{\|u\|_2} \\ &= \sqrt{L} \cdot \sigma_L(M_j^+). \end{aligned}$$

Similarly, we can obtain $\sigma_L(\overline{P_j} M_j^+) \leq \sqrt{L} \cdot \sigma_L(P_j)$ by considering the transposed matrix $\overline{M_j^+} P_j$. □