Section 1.2 - Markov Bases of Hierarchical Models

Leonardo Saud Maia Leite

February 19, 2021

Recall that, given $r_1, r_2, \ldots, r_m \in \mathbb{N}$, $R = \prod_{i=1}^{m} [r_i]$ and $A \in \mathbb{Z}^{d \times R}$, we define

$$\mathcal{M}_A := \{ p \in \mathrm{int}(\Delta_{R-1}) : \log(p) \in \mathrm{image}(A^t) \}.$$

Here, the sum of the entries in each column is constant.

Exp. 1 (Independence): Recall that an $r \times c$ probability table $p = (p_{ij})$ is in $\mathcal{M}_{X \perp\!\!\!\perp Y} \Leftrightarrow$

$p_{ij} = p_{i+} p_{+j}$ for all i, j. If $p_{ij} > 0$ for all i, j, then $\log(p_{ij}) = \log(p_{i+}) + \log(p_{+j})$.

When $r=2$ and $c=3$, then $\log(p)$ is in the row span of the matrix

$$A = \begin{array}{c} \begin{array}{cccccc} 11 & 12 & 13 & 21 & 22 & 23 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

$\hookrightarrow$ in general, A is an $(r+c)\times rc$ matrix

Now, we will compute $\ker_{\mathbb{Z}}(A)$. Let u be a 2×3 matrix written in "vectorized" format, that is,

$$u = (u_{11}\ u_{12}\ u_{13}\ u_{21}\ u_{22}\ u_{23})^t.$$

Then,

$$Au = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \end{pmatrix} = \begin{pmatrix} u_{1+} \\ u_{2+} \\ u_{+1} \\ u_{+2} \\ u_{+3} \end{pmatrix}.$$

For arbitrary r and c, we have

$$Au = \begin{pmatrix} u_{\circ +} \\ u_{+ \circ} \end{pmatrix}.$$

Hence,

$$\ker_{\mathbb{Z}}(A) = \left\{ u \in \mathbb{Z}^{r \times c} : \sum_{k=1}^{r} u_{kj} = 0 \quad \forall j = 1, \dots, c \right.$$

$$\left. \text{and } \sum_{k=1}^{c} u_{ik} = 0 \quad \forall i = 1, \dots, r \right\}.$$

Let e_{ij} be the table with 1 in the (i,j)

basic move position and 0 elsewhere. If u is a vector or a matrix, we define the one-norm of u as $\|u\|_1 := \sum_{i=1}^{R} |u_i|$.

The following proposition presents a basis for $\mathcal{M}_{X \amalg Y}$.

Proposition 2: The unique minimal Markov basis for $\mathcal{M}_{X \perp\!\!\!\perp Y}$ consists of the following $2 \cdot \binom{r}{2}\binom{c}{2}$ moves, each having one norm 4:

$$B = \{ \pm (e_{ij} + e_{kl} - e_{il} - e_{kj}) : 1 \leq i < k \leq r, 1 \leq j < l \leq c \}.$$

Proof: Recall that a basis for a model $\mathcal{M}_A$ is a subset $\mathcal{B} \subset \ker_{\mathbb{Z}}(A)$ such that, for all $w \in \mathcal{T}(n)$ and all pairs $u, v \in \mathcal{F}(w)$, there exists a sequence $w_1, \ldots, w_L \in \mathcal{B}$ such that

$$v = u + \sum_{k=1}^{L} w_k \quad \text{and} \quad u + \sum_{k=1}^{l} w_k \geq 0 \quad \text{for all } l = 1, \ldots, L.$$

Proof: Recall that a basis for a model $\mathcal{M}_A$ is a subset $\mathcal{B} \subset \ker_{\mathbb{Z}}(A)$ such that, for all $w \in \mathcal{T}(n)$ and all pairs $u, v \in \mathcal{F}(w)$, there exists a sequence $w_1, \ldots, w_L \in \mathcal{B}$ such that

$$\boxed{\begin{aligned} \mathcal{T}(n) &= \{w \in \mathbb{N}^R : \textstyle\sum_{i \in R} w_i = n\} \\ \mathcal{F}(w) &= \{v \in \mathbb{N}^R : Av = Aw\} \end{aligned}}$$

$$v = u + \sum_{k=1}^{L} w_k \quad \text{and} \quad u + \sum_{k=1}^{l} w_k \geq 0 \quad \text{for all } l = 1, \ldots, L.$$

Proof: Recall that a basis for a model $\mathcal{M}_A$ is a subset $\mathcal{B} \subset \ker_{\mathbb{Z}}(A)$ such that, for all $w \in \mathcal{T}(n)$ and all pairs $u, v \in \mathcal{F}(w)$, there exists a sequence $w_1, \dots, w_L \in \mathcal{B}$ such that

$$\boxed{\begin{array}{l} \mathcal{T}(n) = \{ w \in \mathbb{N}^R : \sum_{i \in R} w_i = n \} \\ \mathcal{F}(w) = \{ v \in \mathbb{N}^R : Av = Aw \} \end{array}}$$

$$v = u + \sum_{k=1}^{L} w_k \quad \text{and} \quad u + \sum_{k=1}^{l} w_k \geq 0 \quad \text{for all } l = 1, \dots, L.$$

By the characterization of $\ker_{\mathbb{Z}}(X \perp\!\!\!\perp Y)$, it is clear that $\mathcal{B} \subset \ker_{\mathbb{Z}}(X \perp\!\!\!\perp Y)$.

Now, let $u, v \in \mathcal{F}(w)$ for some $w \in \mathcal{T}(n)$, $u \neq v$. Then, u and v are non-negative integral tables that have the same row and column sums. We will prove that there is $b \in B$ such that $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$.

Now, let $u, v \in \mathcal{F}(w)$ for some $w \in \mathcal{T}(n)$, $u \neq v$. Then, u and v are non-negative integral tables that have the same row and column sums. We will prove that there is $b \in B$ such that $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$. $\longrightarrow$ repeat this a finite number of times to get $\|u + b_1 + \dots + b_L - v\|_1 = 0 \Rightarrow v = u + b_1 + \dots + b_L$

Now, let $u, v \in \mathcal{F}(w)$ for some $w \in \mathcal{T}(n)$, $u \neq v$. Then, u and v are non-negative integral tables that have the same row and column sums. We will prove that there is $b \in \mathcal{B}$ such that $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$. $\rightarrow$ repeat this a finite number of times to get $\|u + b_1 + \dots + b_L - v\|_1 = 0 \Rightarrow v = u + b_1 + \dots + b_L$

Since $u \neq v$ and $Au - Av = 0$, then there is at least one positive entry in $u - v$. W.l.o.g., $u_{11} - v_{11} > 0$.

Now, let $u, v \in \mathcal{F}(w)$ for some $w \in \mathcal{T}(n)$, $u \neq v$. Then, u and v are non-negative integral tables that have the same row and column sums. We will prove that there is $b \in \mathcal{B}$ such that $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$. $\longrightarrow$ repeat this a finite number of times to get $\|u + b_1 + \dots + b_L - v\|_1 = 0 \Rightarrow v = u + b_1 + \dots + b_L$

Since $u \neq v$ and $Au - Av = 0$, then there is at least one positive entry in $u - v$. W.l.o.g., $u_{11} - v_{11} > 0$. Since $u - v \in \operatorname{Ker}_{\mathbb{Z}}(X \amalg Y)$, then there is an entry in the first row of $u - v$ that is negative, say $u_{12} - v_{12} < 0$. Similarly, $u_{22} - v_{22} > 0$.

Now, let $u, v \in \mathcal{F}(w)$ for some $w \in \mathcal{T}(n)$, $u \neq v$. Then, u and v are non-negative integral tables that have the same row and column sums. We will prove that there is $b \in \mathcal{B}$ such that $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$. $\rightarrow$ repeat this a finite number of times to get $\|u + b_1 + \ldots + b_L - v\|_1 = 0 \Rightarrow v = u + b_1 + \ldots + b_L$

Since $u \neq v$ and $Au - Av = 0$, then there is at least one positive entry in $u - v$. W.l.o.g., $u_{11} - v_{11} > 0$. Since $u - v \in \operatorname{Ker}_{\mathbb{Z}}(X \perp\!\!\!\perp Y)$, then there is an entry in the first row of $u - v$ that is negative, say $u_{12} - v_{12} < 0$. Similarly, $u_{22} - v_{22} > 0$. So, if we take $b = e_{12} + e_{21} - e_{11} - e_{22}$, then $u + b \geq 0$ and $\|u - v\|_1 > \|u + b - v\|_1$. $\boxtimes$

The tableau notation for a move $e_{ij} + e_{kl} - e_{il} - e_{kj}$ in the Markov basis of $\mathcal{M}_{X \perp\!\!\!\perp Y}$ is

$$\begin{bmatrix} i & j \\ k & l \end{bmatrix} - \begin{bmatrix} i & l \\ k & j \end{bmatrix}.$$

Definition 3: A simplicial complex is a set $\Gamma \subseteq 2^{[m]}$ such that $F \in \Gamma$ and $S \subset F$ implies that $S \in \Gamma$. The elements of Γ are called faces and the inclusion-maximal faces are called facets.

Notation: $\Gamma = [12][13]$ is the bracket notation for the

facets of Γ

simplicial complex $\Gamma = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}\}$.

Hierarchical Log-Linear Models

Hierarchical Log-Linear Models

Recall that, if $\log(p) \in \text{rowspan}(A)$, then $\log(p) = A^t\alpha$ for some $\alpha \in \mathbb{R}^d$, which implies that $p = \exp(A^t\alpha)$.

Hierarchical Log-Linear Models

Recall that, if $\log(p) \in$ rowspan(A), then $\log(p) = A^t\alpha$ for some $\alpha \in \mathbb{R}^d$, which implies that $p = \exp(A^t\alpha)$.

Introducing a normalizing constant, we get

$$p = \frac{1}{Z(\alpha)} \exp(A^t\alpha).$$

Hierarchical Log-Linear Models

Recall that, if $\log(p) \in \text{rowspan}(A)$, then $\log(p) = A^t\alpha$ for some $\alpha \in \mathbb{R}^d$, which implies that $p = \exp(A^t\alpha)$.

Introducing a normalizing constant, we get

$$p = \frac{1}{Z(\alpha)} \exp(A^t \alpha).$$

Setting $\theta_i = \exp(\alpha_i)$, we have

$$p_j = P(X=j) = \frac{1}{Z(\theta)} \prod_{i=1}^{d} \theta_i^{a_{ij}}.$$

$A = (a_{ij})$

Notation: If $i = (i_1, \ldots, i_m) \in R$ and $F = \{f_1 < f_2 < \ldots\} \subset [m]$, then $i_F = (i_{f_1}, i_{f_2}, \ldots)$. For each subset $F \subseteq [m]$, the random vector $X_F = (X_f)_{f \in F}$ has the state space $R_F = \prod_{f \in F} [r_f]$.

Definition 4: Let $\Gamma \subseteq 2^{[m]}$ be a S.C. and $r_1, \ldots, r_m \in \mathbb{N}$.
For each facet $F \in \Gamma$, we introduce a set of $\# R_F$ pos. param. $\theta^{(F)}_{i_F}$. The hierarchical log-linear model associated with Γ is the set of all probability distributions

$$\mathcal{M}_\Gamma = \left\{ p \in \Delta_{R-1} : p_i = \frac{1}{Z(\theta)} \prod_{F,\ \text{facet of } \Gamma} \theta^{(F)}_{i_F} \text{ for all } i \in R \right\},$$

where $Z(\theta) = \sum_{i \in R} \prod_{F,\ \text{facet of } \Gamma} \theta^{(F)}_{i_F}$.

Example 5 (Independence): Let $\Gamma = [1][2]$ and let $R = [r_1] \times [r_2]$ for any $r_1, r_2 \in \mathbb{N}$. Then, the hierarchical model of Γ is the set of all positive prob. matrices $(p_{i_1 i_2})$ such that $p_{i_1, i_2} = \frac{1}{Z(\theta)} \theta^{(1)}_{i_1} \theta^{(2)}_{i_2}$, where $\theta^{(j)}$

is in $(0, \infty)^{r_j}$ for $j = 1, 2$. Hence, $\mathcal{M}_\Gamma$ is the model of all positive rank 1 matrices and it is the positive part of $\mathcal{M}_{X \perp\!\!\!\perp Y}$.

By construction, given a simplicial complex Γ, there is a matrix A_Γ that realizes the model $\mathcal{M}_\Gamma$ in the form $\mathcal{M}_{A_\Gamma}$.

By construction, given a simplicial complex Γ, there is a matrix A_Γ that realize the model $\mathcal{M}_\Gamma$ in the form $\mathcal{M}_{A_\Gamma}$.

$\boxed{\text{Markov bases of hierarchical models}}$

By construction, given a simplicial complex Γ, there is a matrix A_Γ that realize the model $\mathcal{M}_\Gamma$ in the form $\mathcal{M}_{A_\Gamma}$. $\boxed{\text{Markov bases of hierarchical models}}$

Notation: Let $u \in \mathbb{N}^R$ be an $r_1 \times \cdots \times r_m$ contingency table.

For any subset $F = \{f_1 < f_2 < \cdots\} \subseteq [m]$, let $u|_F$ be the $r_{f_1} \times r_{f_2} \times \cdots$ marginal table such that

$\hookrightarrow$ F-marginal of u

$$(u|_F)_{i_F} = \sum_{j \in R_{[m] \setminus F}} u_{i_F, j}.$$

Proposition 6: Let $\Gamma = [F_1][F_2]\cdots$ be a simplicial complex. The matrix A_Γ represents the linear transformation

$$u \mapsto (u|_{F_1}, u|_{F_2}, \dots),$$

and the Γ-marginals are minimal sufficient statistics of the hierarchical model $\mathcal{M}_\Gamma$.

Example 7: Consider $\Gamma = [12][14][23]$ and $r_1 = r_2 = r_3 = r_4 = 2$. Then, A_Γ is construc. as follows:

$$
\begin{array}{l}
 \quad 1111 \quad 1112 \quad 1121 \quad 1122 \quad 1211 \quad 1212 \quad 1221 \quad 1222 \cdots \\
([12], 11) \\
([12], 12) \\
([12], 21) \\
([12], 22) \\
([14], 11) \\
([14], 12) \\
([14], 21) \\
([14], 22) \\
\vdots
\end{array}
\left(\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\right.
$$

Example 8 : Returning to Example 5, for $\Gamma = [1][2]$, the minimal sufficient statistics are the row and column sums of $u \in \mathbb{N}^{r_1 \times r_2}$, that is, they are the vectors

$$A_{[1][2]}\, u = \begin{pmatrix} u_{\bullet +} \\ u_{+ \bullet} \end{pmatrix}.$$

Example 8: Returning to Example 5, for $\Gamma = [1][2]$, the minimal sufficient statistics are the row and column sums of $u \in \mathbb{N}^{r_1 \times r_2}$, that is, they are the vectors

$\boxed{R_{[2]} = [r_2]}$

$$A_{[1][2]} u = \begin{pmatrix} u_{\bullet +} \\ u_{+ \bullet} \end{pmatrix}.$$

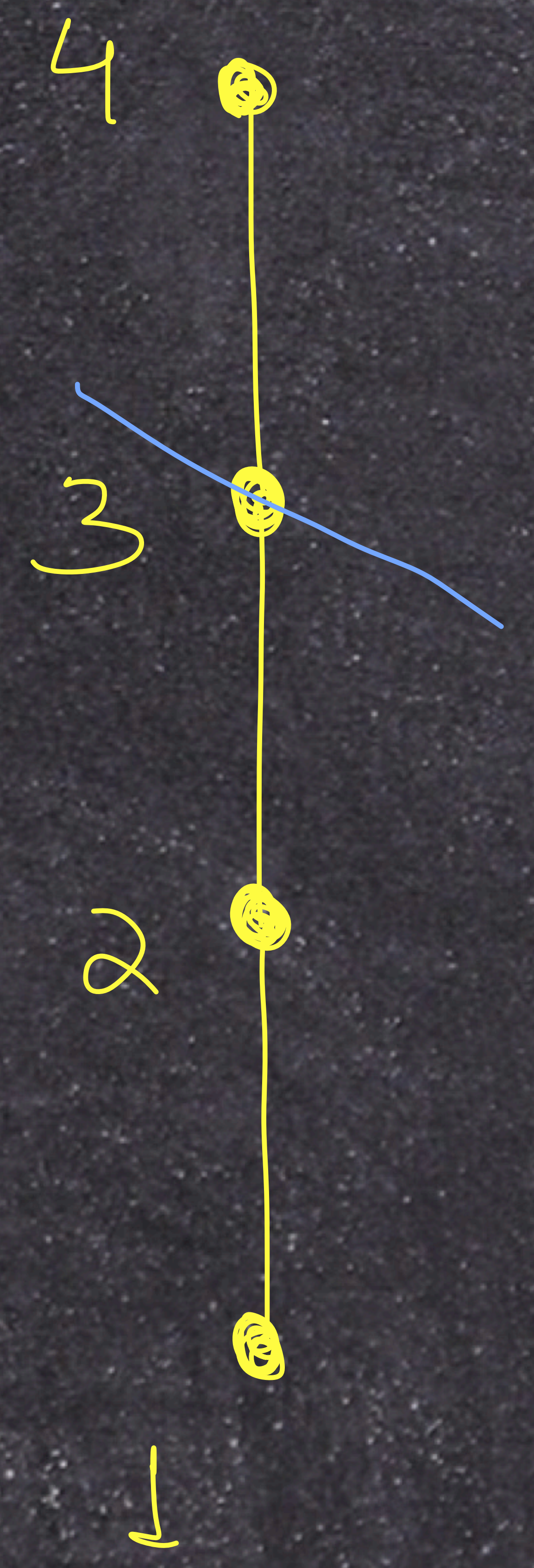
But $(u|_1)_{i_1} = \sum_{j \in R_{[2]}} u_{i_1, j}$ and $(u|_2)_{i_2} = \sum_{j \in R_{[1]}} u_{j, i_2}$.

So, $A_{[1][2]} u = (u|_1, u|_2)$.

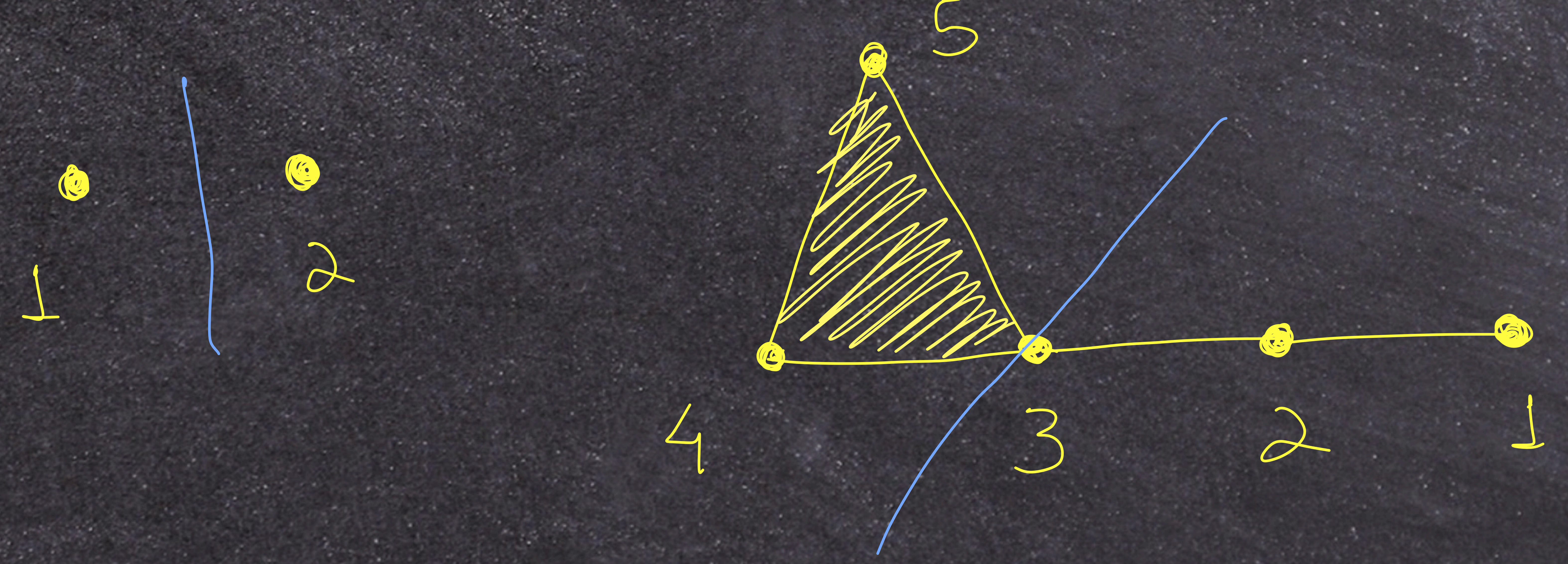
For a simplicial complex Γ, we define the ground set of Γ as $G(\Gamma) = \bigcup_{S \in \Gamma} S$.

Definition 9: A simplicial complex is reducible, with reducible decomposition (Γ_1, S, Γ_2) and separator $S \subset G(\Gamma)$, if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$, $\Gamma_1 \cap \Gamma_2 = 2^S$ and $\Gamma_1, \Gamma_2 \not\subseteq 2^S$.
A simplicial complex is decomposable if it is reducible and Γ_1 and Γ_2 are decomposable or simplices (that is, of the form 2^R for some $R \subseteq [m]$).

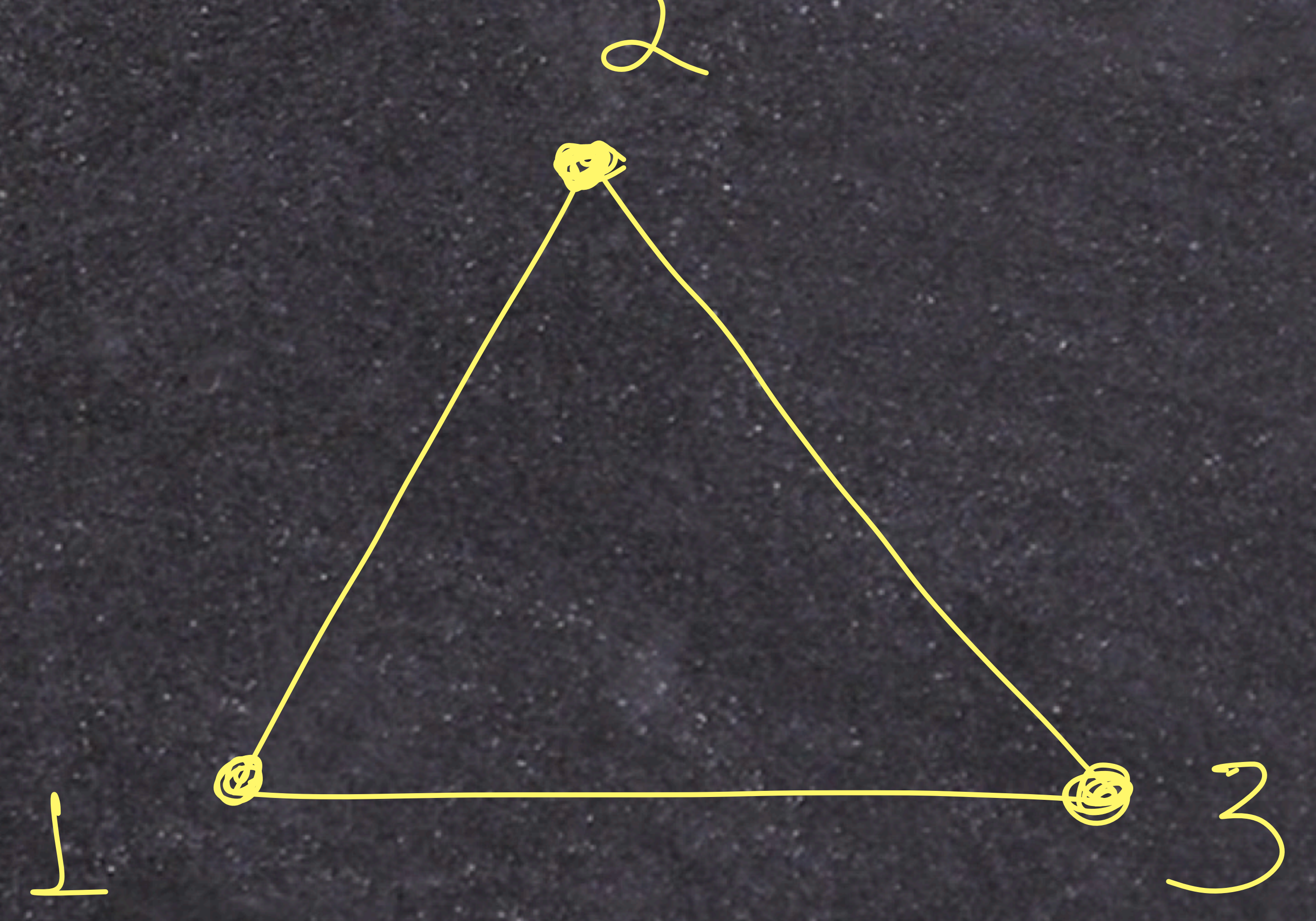
Example 10: Consider $\Gamma = [12][23][34]$. Γ is reducible with reducible decomposition $([12][23], \{3\}, [34])$. It is also decomposable, since Γ_1 is decomposable.



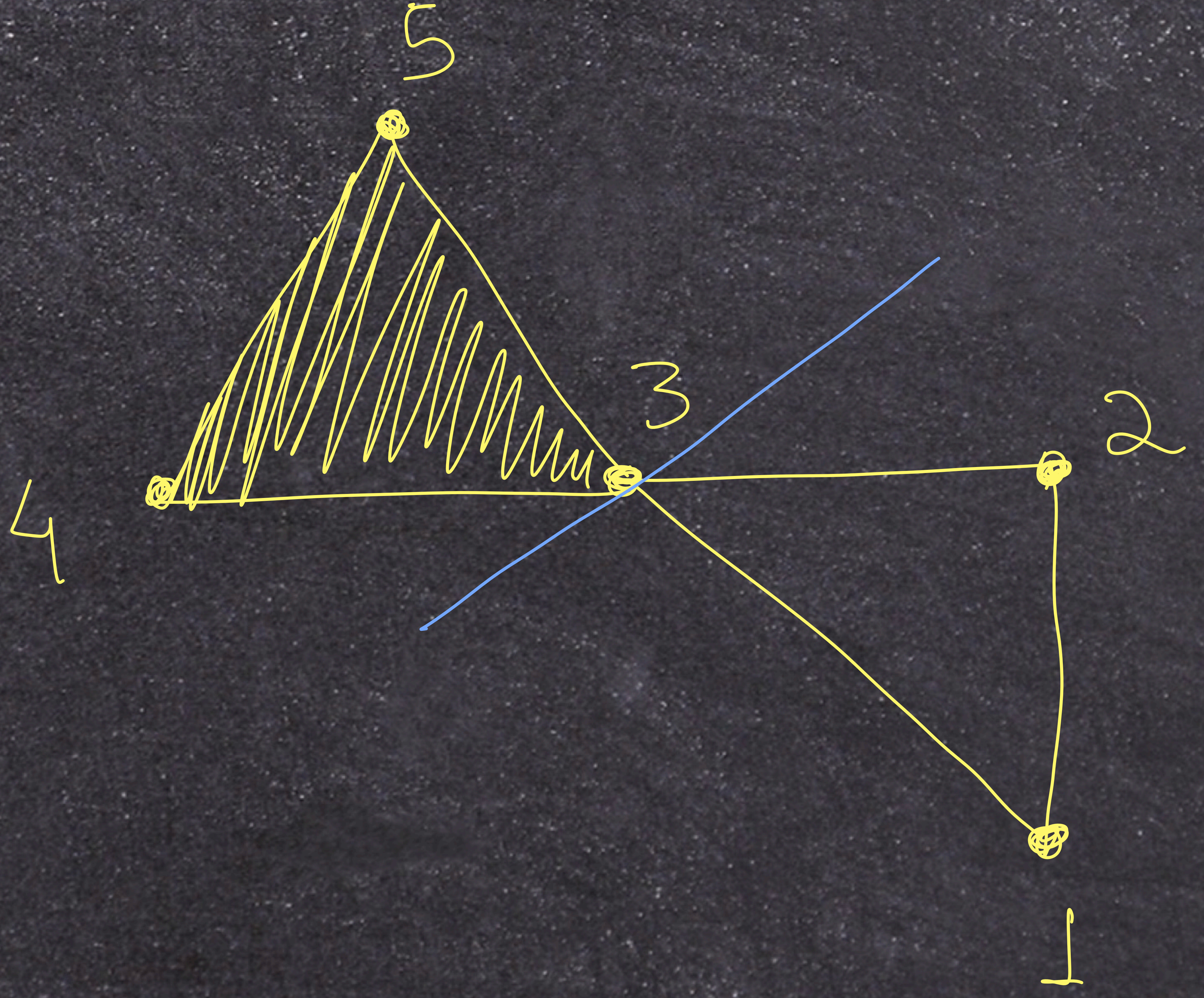
Example 11: [1] [2] and [12][23][345] are also dec.



Example 12: [12][13][23] is not reducible.



Example 13: [12] [13] [23] [345] is reducible with red. decomposition ([12] [13] [23], 3, [345]), but not decomp.



Lemma 14: If Γ is a reducible S.C. with reducible decomposition (Γ_1, S, Γ_2), then the following set of moves belongs to $\mathrm{Ker}_{\mathbb{Z}}(A_\Gamma)$:

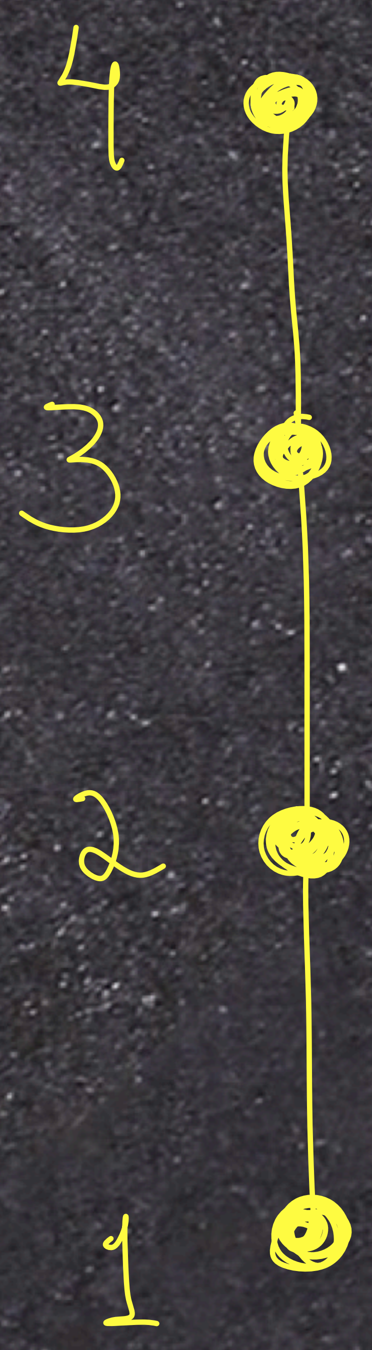
$$D(\Gamma_1,\Gamma_2)=\left\{\begin{bmatrix} i & j & k \\ i' & j & k' \end{bmatrix}-\begin{bmatrix} i & j & k' \\ i' & j & k \end{bmatrix} : i,i' \in R_{G(\Gamma_1)\setminus S},\ j\in R_S,\ k,k' \in R_{G(\Gamma_2)\setminus S}\right\}.$$

Theorem 15 (Markov bases of decomposable models): If Γ is a decomposable S. C., then the set of moves

$$\mathcal{B} = \bigcup_{(\Gamma_1, S, \Gamma_2)} D(\Gamma_1, \Gamma_2),$$

with the union over all reducible decompositions of Γ, is a Markov basis for A_Γ.

Example 16: Consider $\Gamma = [12][23][34]$. Γ has two distinct reducible decompositions with minimal separator: $([12], \{2\}, [23][34])$ and $([12][23], \{3\}, [34])$. Therefore, the Markov basis of Γ is given by



$$D([12],[23][34]) \cup D([12][23],[34]),$$

which, in tableau notation, is given by

$$\begin{bmatrix} i_1 & j & i_3 & i_4 \\ i_1' & j & i_3' & i_4' \end{bmatrix} - \begin{bmatrix} i_1 & j & i_3' & i_4' \\ i_1' & j & i_3 & i_4 \end{bmatrix} \text{ and } \begin{bmatrix} i_1 & i_2 & j & i_4 \\ i_1' & i_2' & j & i_4' \end{bmatrix} - \begin{bmatrix} i_1 & i_2 & j & i_4' \\ i_1' & i_2' & j & i_4 \end{bmatrix}.$$

Any questions?