

Explain the basics of Determinantal Point Process

Richard Xu

January 11, 2022

1 What is DPP?

Most of this note is based on the original DPP paper <https://arxiv.org/abs/1207.6083>. The original paper is very detailed and well written. However, there may be some points that need further clarification, especially for students lacking linear algebra skill. Therefore, I hope to explain them in a slightly simpler way (hopefully). Please read the original paper for more details.

1.1 definition of marginal DPP distribution

Start with a **marginal** distribution:

$$\Pr(A \subseteq \mathbf{Y}) = \det(K_A) \quad (1)$$

An example: given $\Omega = \{1, 2, 3, 4, 5\}$, $A = \{1, 2, 3\}$ and $\mathbf{Y} \in \Omega$

$$\begin{aligned} \Pr(A \subseteq \mathbf{Y}) &= \Pr(\{1, 2, 3\} \subseteq \mathbf{Y}) \\ &\equiv \Pr_K(y_1 = 1, y_2 = 1, y_3 = 1) \\ &= \sum_{t_4=0}^1 \sum_{t_5=0}^1 \Pr(y_1 = 1, y_2 = 1, y_3 = 1, y_4 = t_4, y_5 = t_5) \\ &= \det(K_A) \end{aligned} \quad (2)$$

1.2 Something about marginal distribution

1. $\Pr(A \subseteq \mathbf{Y})$ is marginal, so $\Pr(A_1 \subseteq \mathbf{Y}) + \Pr(A_2 \subseteq \mathbf{Y}) + \dots$ don't need to add to 1, i.e., it may be possible that: $\Pr(A_1 \subseteq \mathbf{Y}) + \Pr(A_2 \subseteq \mathbf{Y}) > 1$
2. $\Pr(\emptyset \subseteq \mathbf{Y}) = \det(K_\emptyset) = 1$ This is obvious, as any \mathbf{Y} is a superset of \emptyset .
3. $\Pr(i \subseteq \mathbf{Y}) = \det(K_{ii}) = K_{ii}$
4. however, its property is best determined from two elements case:

$$\begin{aligned} \Pr(i, j \in \mathbf{Y}) &= \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} \\ &= K_{ii}K_{jj} - K_{ij}K_{ji} \\ &= \Pr(i \subseteq \mathbf{Y})\Pr(j \subseteq \mathbf{Y}) - K_{ij}^2 \end{aligned} \quad (3)$$

By convention, off-diagonal elements determine negative correlations between pairs.

Large absolute values of $K_{i,j}$ imply that the probability that i^{th} and j^{th} elements are both selected tend to have **low** density.

1.2.1 Example of K

Any $K, 0 \preceq K \preceq I$ defines a DPP.

If $A \preceq B$, that is, $B - A$ is positive semi-definite.

1.2.2 where K does not define DPP

example $K = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ does **not** define DPP, we check if $K \preceq I$?

$$\begin{aligned} I - \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} &= \begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix} \\ \implies \bar{\lambda}(K) &= [-0.5, 0.5]^\top \end{aligned} \quad (4)$$

Another way to see the above is incorrect, where we let $\Omega = \{1, 2\}$:

$$\begin{aligned} \Pr(\{1\} \subseteq \mathbf{Y}) &\equiv \Pr((\mathbf{Y} = \{1\}) \cup (\mathbf{Y} = \{1, 2\})) \\ &= \det(K_1) = 1 \end{aligned} \quad (5)$$

$$\begin{aligned} \Pr(\{2\} \subseteq \mathbf{Y}) &\equiv \Pr((\mathbf{Y} = \{2\}) \cup (\mathbf{Y} = \{1, 2\})) \\ &= \det(K_2) = 1 \end{aligned} \quad (6)$$

note LHS uses \subseteq and RHS uses $=$. However:

$$\begin{aligned} \Pr(\{1, 2\} \subseteq \mathbf{Y}) &\equiv \Pr(\mathbf{Y} = \{1, 2\}) \\ &= \det(K_{\{1, 2\}}) = 0.75 \end{aligned} \quad (7)$$

1. The first two equation says $\{1\}$ and $\{2\}$ must be included
2. The third equation says both may NOT always be included

1.2.3 Example of K define DPP

example $K = \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.4 \end{bmatrix}$ **does** define DPP:

$$\begin{aligned} I - \begin{pmatrix} 0.3 & -0.1 \\ -0.1 & 0.4 \end{pmatrix} &= \begin{pmatrix} 0.7 & 0.1 \\ 0.1 & 0.6 \end{pmatrix} \\ \implies \bar{\lambda}(K) &= [0.5382, 0.7618]^\top \end{aligned} \quad (8)$$

$$\begin{aligned} \Pr(\{1\} \subseteq \mathbf{Y}) &\equiv \Pr((\mathbf{Y} = \{1\}) \cup (\mathbf{Y} = \{1, 2\})) \\ &= \det(K_1) = 0.3 \end{aligned} \quad (9)$$

$$\begin{aligned} \Pr(\{2\} \subseteq \mathbf{Y}) &\equiv \Pr((\mathbf{Y} = \{2\}) \cup (\mathbf{Y} = \{1, 2\})) \\ &= \det(K_2) = 0.4 \end{aligned} \quad (10)$$

$$\begin{aligned} \Pr(\{1, 2\} \subseteq \mathbf{Y}) &\equiv \Pr(\mathbf{Y} = \{1, 2\}) \\ &= \det(K_{\{1, 2\}}) = 0.11 \end{aligned} \quad (11)$$

the event:

$$\begin{aligned}\Pr(\{1, 2\} \subseteq \mathbf{Y}) &\equiv \Pr(\{1, 2\} = \mathbf{Y}) \\ &= \Pr(\{1\} \subseteq \mathbf{Y}) \cap (\{2\} \subseteq \mathbf{Y})\end{aligned}\tag{12}$$

$$\begin{aligned}\Pr((\{1\} = \mathbf{Y}) \cup (\{2\} = \mathbf{Y})) &= \Pr(\{1\} \subseteq \mathbf{Y}) + \Pr(\{2\} \subseteq \mathbf{Y}) - \Pr(\{1, 2\} \subseteq \mathbf{Y}) \\ &= 0.3 + 0.4 - 0.11 \\ &= 0.59\end{aligned}\tag{13}$$

what about the probability of selecting **exactly** the \emptyset ?

$$\begin{aligned}\Pr(\mathbf{Y} = \emptyset) &\equiv 1 - \Pr((\{1\} = \mathbf{Y}) \cup (\{2\} = \mathbf{Y})) \\ &= 0.41\end{aligned}\tag{14}$$

2 L-Ensembles

Marginal distributions does **not** define probabilities in terms of a **particular** set directly, i.e., instead of having $\Pr(\mathbf{Y} \subseteq Y)$, we want $\Pr(\mathbf{Y} = Y)$:

$$\Pr_L(\mathbf{Y} = Y) \propto \det(L_Y)\tag{15}$$

L must be positive semidefinite.

Only a statement of proportionality, eigenvalues of L is **not** < 1

2.1 Geometry interpretation

$$\begin{aligned}X &= [x_1 \quad x_2 \quad \dots \quad x_n] \implies \\ L(x_1, \dots, x_n) &= X^\top X = \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \dots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_2, x_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \dots & \langle x_n, x_n \rangle \end{pmatrix}\end{aligned}\tag{16}$$

Gram determinant is the square of the volume of the parallelotope formed by the vectors
vectors are linearly independent if and only if the Gram determinant is nonzero

$$\Pr_L(Y) \propto \det(L_Y) = \text{Vol}^2(\{x_i\}_{i \in Y})$$

2.2 Proof for the Geometry interpretation

2.2.1 in 1-element case

$$\text{Vol}^2(\mathbf{u}_1) = \mathbf{u}_1^\top \mathbf{u}_1, \text{ i.e., length square of a line}$$

2.2.2 in k-element case

$$\text{Vol}^2(\mathbf{u}_1 \dots \mathbf{u}_k, \mathbf{u}_{k+1}) = \text{Vol}^2(\mathbf{u}_1, \dots, \mathbf{u}_k) \|\tilde{\mathbf{u}}_{k+1}\|^2\tag{17}$$

$\tilde{\mathbf{u}}_{k+1}$ is the orthogonal projection of \mathbf{u}_{k+1} onto $\text{span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$:

Let $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is an $n \times k$ matrix \mathbf{Y} :

Then there exists a vector $\mathbf{c} \in \mathbb{R}^k$ such that:

$\mathbf{u}_{k+1} = \mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1}$ split \mathbf{u}_{k+1} into \parallel and \perp components regarding span $(\mathbf{u}_1, \dots, \mathbf{u}_k)$

$$= \underbrace{\begin{bmatrix} | & \vdots & | \\ \mathbf{u}_1 & \vdots & \mathbf{u}_k \\ | & \vdots & | \end{bmatrix}}_{\mathbf{Y}} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} + \tilde{\mathbf{u}}_{k+1} \quad \text{or } \mathbf{u}_{k+1} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 \dots c_k \mathbf{u}_k + \tilde{\mathbf{u}}_{k+1} \quad (18)$$

extending $\mathbf{Y} \rightarrow \mathbf{X}$ by adding one more column \mathbf{u}_{k+1} :

$$\begin{aligned} \mathbf{X} &= [\mathbf{Y} \quad \mathbf{u}_{k+1}] = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_k \quad \mathbf{u}_{k+1}] = [\mathbf{Y} \quad \mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1}] \\ \Rightarrow \mathbf{X}^\top \mathbf{X} &= \begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{Y}^\top \mathbf{u}_{k+1} \\ \mathbf{u}_{k+1}^\top \mathbf{Y} & \mathbf{u}_{k+1}^\top \mathbf{u}_{k+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{Y}^\top (\mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1}) \\ (\mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1})^\top \mathbf{Y} & (\mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1})^\top (\mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1}) \end{bmatrix} \quad \text{using } \mathbf{u}_{k+1} = \mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1} \\ &= \begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} & \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y}\mathbf{c} + \tilde{\mathbf{u}}_{k+1}^\top \tilde{\mathbf{u}}_{k+1} \end{bmatrix} \quad \text{since } \mathbf{Y}^\top \tilde{\mathbf{u}}_{k+1} = \mathbf{0} \\ &= \begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} & \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y}\mathbf{c} + \|\tilde{\mathbf{u}}_{k+1}\|^2 \end{bmatrix} \\ &= \left[\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} \end{bmatrix} \quad \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \|\tilde{\mathbf{u}}_{k+1}\|^2 \end{bmatrix} \right) \right] \end{aligned} \quad (19)$$

$$\begin{aligned} \det([a_1 + b_1, a_2, \dots, a_k]) &= \det([a_1, a_2, \dots, a_k]) + \det([b_1, a_2, \dots, a_k]) \quad \text{using Multi-linearity} \\ \Rightarrow \det(\mathbf{X}^\top \mathbf{X}) &= \det \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} \end{bmatrix} \quad \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \|\tilde{\mathbf{u}}_{k+1}\|^2 \end{bmatrix} \right) \right) \\ &= \det \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} & \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y}\mathbf{c} \end{bmatrix} \right) + \det \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{0} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} & \|\tilde{\mathbf{u}}_{k+1}\|^2 \end{bmatrix} \right) \\ &= \mathbf{0} + \det \left(\begin{bmatrix} \mathbf{Y}^\top \mathbf{Y} & \mathbf{0} \\ \mathbf{c}^\top \mathbf{Y}^\top \mathbf{Y} & \|\tilde{\mathbf{u}}_{k+1}\|^2 \end{bmatrix} \right) \\ &= \det \left([\mathbf{Y}^\top \mathbf{Y}] \right) \|\tilde{\mathbf{u}}_{k+1}\|^2 \\ &= \det \left([\mathbf{Y}^\top \mathbf{Y}] \right) \text{Vol}^2(\tilde{\mathbf{u}}_{k+1}) \end{aligned} \quad (20)$$

2.3 Normalization constant in L-Ensembles

without proof, stating the **Theorem** says:

Theorem 1

$$\sum_{A \subseteq Y \subseteq \Omega} \det(L_Y) = \det(L + \mathbf{I}_{\bar{A}}) \quad (21)$$

2.3.1 illustrate through an example

from this, it can be easily understood by multilinear rule:

$$\begin{aligned}
L &= \begin{pmatrix} 3.0 & 1.0 \\ 1.5 & 1.2 \end{pmatrix} \\
\det(L + \mathbf{I}) &= \det \left(\begin{bmatrix} 3.0+1 & 1.0+0 \\ 1.5+0 & 1.2+1 \end{bmatrix} \right) \\
&= \det \left(\begin{bmatrix} 3.0 & 1.0+0 \\ 1.5 & 1.2+1 \end{bmatrix} \right) + \det \left(\begin{bmatrix} 1 & 1.0+0 \\ 0 & 1.2+1 \end{bmatrix} \right) \\
&= \det \left(\begin{bmatrix} 3.0 & 1.0 \\ 1.5 & 1.2 \end{bmatrix} \right) + \det \left(\begin{bmatrix} 3.0 & 0 \\ 1.5 & 1 \end{bmatrix} \right) + \det \left(\begin{bmatrix} 1 & 1.0 \\ 0 & 1.2 \end{bmatrix} \right) + \det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\
&= \det \left(\begin{bmatrix} 3.0 & 1.0 \\ 1.5 & 1.2 \end{bmatrix} \right) + \det([3.0]) + \det([1.2]) + \det(\mathbf{I}) \\
&= \underbrace{\det(L)}_{\{1,2\}} + \underbrace{\det([3.0])}_{\{1\}} + \underbrace{\det([1.2])}_{\{2\}} + \underbrace{1}_{\emptyset}
\end{aligned} \tag{22}$$

Note that unless we are interested to compute $\sum_{A \subseteq Y} L_Y$ where $A = \emptyset$, we will **not** have the term $\det(\mathbf{I}) = 1$. This is not surprising, as $\sum_{A \subseteq Y} L_Y$ terms do not contain \emptyset . from determinant computation point of view, multilinear rule will not result to a full I , there will be some zeros.

1. those include $\{1\}$

$$\begin{aligned}
\sum_{\{1\} \subseteq Y} \det(L_Y) &= \underbrace{\det(L)}_{\{1,2\}} + \underbrace{\det([3.0])}_{\{1\}} \quad \text{by derivation} \\
&= \det(L + \mathbf{I}_{\{\bar{1}\}}) \quad \text{by theorem 1} \\
&= \det \left(\begin{bmatrix} 3.0+0 & 1.0+0 \\ 1.5+0 & 1.2+1 \end{bmatrix} \right)
\end{aligned} \tag{23}$$

2. those include $\{2\}$

$$\begin{aligned}
\sum_{\{2\} \subseteq Y} \det(L_Y) &= \underbrace{\det(L)}_{\{1,2\}} + \underbrace{\det([1.2])}_{\{2\}} \quad \text{by derivation} \\
&= \det(L + \mathbf{I}_{\{\bar{2}\}}) \quad \text{by theorem 1} \\
&= \det \left(\begin{bmatrix} 3.0+1 & 1.0+0 \\ 1.5+0 & 1.2+0 \end{bmatrix} \right)
\end{aligned} \tag{24}$$

2.3.2 another example

$$\begin{aligned}
L &= \begin{pmatrix} 2.8599 & -0.4936 & -1.8458 \\ -0.4936 & 2.6264 & -1.1437 \\ -1.8458 & -1.1437 & 2.0522 \end{pmatrix} \\
A = \{1, 2\} &\implies \bar{A} = \{3\} \implies \mathbf{I}_{\bar{A}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned} \tag{25}$$

Apply theorem (1):

$$\sum_{\{1,2\} \subseteq Y} \det(L_Y) = \det \left(\begin{bmatrix} 2.8599+0 & -0.4936+0 & -1.8458+0 \\ -0.4936+0 & 2.6264+0 & -1.1437+0 \\ -1.8458+0 & -1.1437+0 & 2.0522+1 \end{bmatrix} \right) \tag{26}$$

2.3.3 let $A = \emptyset$

from Theorem , normalisation constant (or partition function) is: $\bar{\emptyset} = \Omega$:

$$\begin{aligned}
\sum_{\emptyset \subseteq Y \subseteq \Omega} \det(L_Y) &= \sum_{Y \subseteq \Omega} \det(L_Y) \\
&= \det(L + \mathbf{I}_{\bar{\emptyset}}) \\
&= \det(L + \mathbf{I}_{\Omega}) \\
&= \det(L + \mathbf{I})
\end{aligned} \tag{27}$$

2.4 Conversion to Marginal distribution

$$\Pr_L(\mathbf{Y} = Y) \propto \det(L_Y) \implies \Pr_L(\mathbf{Y} = Y) = \frac{\det(L_Y)}{\det(L_Y + I)} \tag{28}$$

An L -ensemble is a DPP, and its marginal kernel is:

$$K = L(L + I)^{-1} = I - (L + I)^{-1} \tag{29}$$

an important identity:

$$L(L + I)^{-1} = I - (L + I)^{-1} \tag{30}$$

for any L where $(L + I)^{-1}$ exist

$$\begin{aligned}
\Pr_L(A \subseteq \mathbf{Y}) &= \frac{\sum_{A \subseteq Y \subseteq \Omega} \det(L_Y)}{\sum_{Y \subseteq \Omega} \det(L_Y)} \\
&= \frac{\det(L + I_{\bar{A}})}{\det(L + I)} \\
&= \det((L + I_{\bar{A}})(L + I)^{-1}) \quad \because \det(A^{-1}) = \frac{1}{\det(A)} \quad \det(AB) = \det(A) \det(B)
\end{aligned} \tag{32}$$

$$\begin{aligned}
\Pr_L(A \subseteq \mathbf{Y}) &= \det((L + I_{\bar{A}})(L + I)^{-1}) \\
&= \det(L(L + I)^{-1} + I_{\bar{A}}(L + I)^{-1}) \quad \text{expand} \\
&= \det(I - (L + I)^{-1} + I_{\bar{A}}(L + I)^{-1}) \quad \because \text{of Eq. (30)} \\
&= \det(I - (I - I_{\bar{A}})(L + I)^{-1}) \quad \text{combine last two terms together} \\
&= \det(I - I_A(L + I)^{-1}) \quad \because I_A = I - I_{\bar{A}} \\
&= \det((I_A + I_{\bar{A}}) - I_A(L + I)^{-1}) \quad \text{expanding } I = I_A + I_{\bar{A}} \\
&= \det(I_{\bar{A}} + \overbrace{I_A - I_A(L + I)^{-1}}) \\
&= \det(I_{\bar{A}} + I_A \underbrace{(I - (L + I)^{-1})}) \quad \because K = I - (L + I)^{-1} \\
&= \det(I_{\bar{A}} + I_A K)
\end{aligned} \tag{33}$$

left multiplication by I_A **zeros out rows** of a matrix except those corresponding to A . We split the marginal kernel matrix K into K_A and $K_{\bar{A}}$:

$$\begin{aligned} K &= \begin{pmatrix} K_{\bar{A}} & K_{\bar{A}A} \\ K_{A\bar{A}} & K_A \end{pmatrix} \\ \implies I_A K &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{|A| \times |A|} \end{pmatrix} \begin{pmatrix} K_{\bar{A}} & K_{\bar{A}A} \\ K_{A\bar{A}} & K_A \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ K_{A\bar{A}} & K_A \end{pmatrix} \end{aligned} \quad (34)$$

Re-organise:

$$\begin{aligned} \Pr_L(A \subseteq \mathbf{Y}) &= \det(I_{\bar{A}} + I_A K) \\ &= \det \left(\begin{bmatrix} I_{|\bar{A}| \times |\bar{A}|} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ K_{A\bar{A}} & K_A \end{bmatrix} \right) \\ &= \det \left(\begin{bmatrix} I_{|\bar{A}| \times |\bar{A}|} & \mathbf{0} \\ K_{A\bar{A}} & K_A \end{bmatrix} \right) \\ &= \det(K_A) \end{aligned} \quad (35)$$

therefore, the conversion formula is:

$$K = L(L + I)^{-1} = I - (L + I)^{-1} \quad (36)$$

2.4.1 Eigen-value conversion

$$K = L(L + I)^{-1} = I - (L + I)^{-1} \quad (37)$$

Properties

$$\begin{aligned} \lambda_n \in \text{eig}(A) &\implies \lambda_n + 1 \in \text{eig}(A + I) \\ &\implies (\lambda_n)^{-1} \in \text{eig}(A^{-1}) \end{aligned} \quad (38)$$

Apply it to $K = I - (L + I)^{-1}$:

$$\begin{aligned} (\lambda_n + 1) \in \text{eig}(L + I) &\implies \frac{1}{\lambda_n + 1} \in \text{eig}((L + I)^{-1}) \\ &\implies 1 - \frac{1}{\lambda_n + 1} \in \text{eig}(I - (L + I)^{-1}) \end{aligned} \quad (39)$$

$$1 - \frac{1}{\lambda_n + 1} = \frac{\lambda_n + 1 - 1}{\lambda_n + 1} = \frac{\lambda_n}{\lambda_n + 1} \quad (40)$$

Therefore,

$$L = \sum_{n=1}^N \lambda_n v_n v_n^\top \implies K = \sum_{n=1}^N \frac{\lambda_n}{\lambda_n + 1} v_n v_n^\top \quad (41)$$

2.4.2 Conversions from K to L

$$K = L(L + I)^{-1} = I - (L + I)^{-1} \quad (42)$$

$$\begin{aligned} K = I - (L + I)^{-1} &\implies I - K = (L + I)^{-1} \\ &\implies (L + I)(I - K) = I \\ &\implies L + I - LK - K = I \\ &\implies L(I - K) = K \\ &\implies L = K(I - K)^{-1} \end{aligned} \quad (43)$$

3 Complement

If \mathbf{Y} is distributed as a DPP with marginal kernel K , then $\Omega - \mathbf{Y}$ is also distributed as a DPP, with marginal kernel $\bar{K} = I - K$:

$$\begin{aligned} \Pr((A \cap \mathbf{Y}) = \emptyset) &= \det(\bar{K}_A) \\ &= \det(I_A - K_A) \end{aligned} \quad (44)$$

For example:

$$K = \begin{pmatrix} 0.4 & 0.1 & -0.1 \\ 0.05 & 0.5 & 0.1 \\ -0.01 & 0.1 & 0.3 \end{pmatrix} \quad A = \{1, 2\} \quad \bar{A} = \{3\} \quad (45)$$

$$\bar{K} = I - K = \begin{pmatrix} 0.6 & -0.1 & 0.1 \\ -0.05 & 0.5 & -0.1 \\ 0.01 & -0.1 & 0.7 \end{pmatrix} \implies \bar{K}_{A=\{1,2\}} = \begin{pmatrix} 0.6 & -0.1 \\ -0.05 & 0.5 \end{pmatrix} \quad (46)$$

It's easy to see that $\bar{K}_A = (I_A - K_A)$

3.0.1 Complement in two point cases

this is just a generalization of Eq.(14):

$$\begin{aligned} \Pr(i, j \notin \mathbf{Y}) &= 1 - \Pr((i \in \mathbf{Y}) \cup (j \in \mathbf{Y})) \\ &= 1 - (\Pr(i \in \mathbf{Y}) + \Pr(j \in \mathbf{Y}) - \Pr(i, j \in \mathbf{Y})) \\ &= 1 - \Pr(i \in \mathbf{Y}) - \Pr(j \in \mathbf{Y}) + \Pr(i, j \in \mathbf{Y}) \\ &\leq 1 - \Pr(i \in \mathbf{Y}) - \Pr(j \in \mathbf{Y}) + \Pr(i \in \mathbf{Y}) \Pr(j \in \mathbf{Y}) \quad \text{from DPP definition: } \Pr(i \in \mathbf{Y}) \Pr(j \in \mathbf{Y}) \geq \Pr(i, j \in \mathbf{Y}) \\ &= 1 - \Pr(i \in \mathbf{Y}) + (1 - \Pr(j \in \mathbf{Y})) - 1 + (1 - \Pr(i \notin \mathbf{Y}))(1 - \Pr(j \notin \mathbf{Y})) \quad \in \rightarrow \notin \\ &= \Pr(i \notin \mathbf{Y}) + \Pr(j \notin \mathbf{Y}) - 1 + (1 - \Pr(i \notin \mathbf{Y}))(1 - \Pr(j \notin \mathbf{Y})) \\ &= \Pr(i \notin \mathbf{Y}) + \Pr(j \notin \mathbf{Y}) - 1 + \underline{1 - \Pr(i \notin \mathbf{Y}) - \Pr(j \notin \mathbf{Y}) + \Pr(i \notin \mathbf{Y}) \Pr(j \notin \mathbf{Y})} \quad \text{expand out} \\ &= \Pr(i \notin \mathbf{Y}) \Pr(j \notin \mathbf{Y}) \end{aligned} \quad (47)$$

Complement of a diversifying process also encourage diversity. (the determinant \bar{K}_A also has the property).

3.0.2 Larger marginal distribution

$$K \preceq K' \implies \det(K_A) \leq \det(K'_A) \quad \forall A \subseteq \Omega \quad (48)$$

DPP defined by K' is “larger” than the one defined by K in the sense that it assigns higher marginal probabilities to every set A .

4 Quality vs Diversity

Think of a Gram matrix, let each column matrix x_i :

$$q_i = \|x_i\|_{L_2} \quad \phi_i = \frac{x_i}{q_i} \implies \|\phi_i\| = 1 \quad (49)$$

$$\text{Let } Q = \begin{bmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & q_n \end{bmatrix} \implies [q_1\phi_1 \quad q_2\phi_2 \quad \dots \quad q_n\phi_n] = \Phi Q$$

$$L(x_1, \dots, x_n) = X^\top X = (\Phi Q)^\top (\Phi Q) = Q^\top \Phi^\top \Phi Q \\ \implies L_{ij} = q_i \phi_i^\top \phi_j q_j \quad (50)$$

$$S_{ij} \equiv \phi_i^\top \phi_j \in [-1, 1] \implies S_{ij} = \frac{L_{ij}}{\sqrt{L_{ii}L_{jj}}}$$

$\Pr_L(\mathbf{Y} = Y)$ can be viewed as the product of four determinants

$$\Pr_L(\mathbf{Y} = Y) \propto \left(\prod_{i \in Y} q_i^2 \right) \det(S_Y) \quad (51)$$

5 Conditional

5.1 $\Pr_L(\mathbf{Y} = B \mid \mathbf{Y} \cap A = \emptyset)$

assuming $B \cap A = \emptyset$, and $B \subseteq \Omega$

$$\begin{aligned} \Pr_L(\mathbf{Y} = B \mid \mathbf{Y} \cap A = \emptyset) &= \frac{\Pr_L((\mathbf{Y} = B) \cap (A \cap \mathbf{Y} = \emptyset))}{\Pr_L(A \cap \mathbf{Y} = \emptyset)} \\ &= \frac{\Pr_L(A \cap \mathbf{Y} = \emptyset \mid \mathbf{Y} = B) \Pr_L(\mathbf{Y} = B)}{\Pr_L(A \cap \mathbf{Y} = \emptyset)} \\ &= \frac{1 \times \Pr_L(\mathbf{Y} = B)}{\Pr_L(A \cap \mathbf{Y} = \emptyset)} \quad \because B \cap A = \emptyset \implies \Pr_L(A \cap \mathbf{Y} = \emptyset \mid \mathbf{Y} = B) = 1 \\ &= \frac{\Pr_L(\mathbf{Y} = B)}{\Pr_L(A \cap \mathbf{Y} = \emptyset)} \\ &= \frac{\frac{\det(L_B)}{\det(L_\Omega + I)}}{\frac{\sum_{B': B' \cap A = \emptyset} \det(L_{B'})}{\det(L_\Omega + I)}} \quad \text{definition of L-Ensembles} \\ &= \frac{\det(L_B)}{\sum_{B': B' \cap A = \emptyset} \det(L_{B'})} \quad \{B' : B' \cap A = \emptyset\} = \bar{A} \\ &= \frac{\det(L_B)}{\det(L_{\bar{A}} + I_{|\bar{A}| \times |\bar{A}|})} \quad \because \Omega \rightarrow \bar{A} \text{ which also include } \emptyset \end{aligned} \quad (52)$$

where $L_{\bar{A}}$ is L indexed by elements in $\Omega \setminus A$.

note that by definition, $I_{|\bar{A}| \times |\bar{A}|} \neq I_{\bar{A}}$

5.2 $\Pr_L(\mathbf{Y} = A \cup B | A \subseteq \mathbf{Y})$

again, assuming $B \cap A = \emptyset$, and $B \subseteq \Omega$

$$\begin{aligned}
\Pr_L(\mathbf{Y} = A \cup B | A \subseteq \mathbf{Y}) &= \frac{\Pr_L((\mathbf{Y} = A \cup B) \cap (A \subseteq \mathbf{Y}))}{\Pr_L(A \subseteq \mathbf{Y})} \\
&= \frac{\underbrace{\Pr_L(A \subseteq \mathbf{Y} | \mathbf{Y} = A \cup B)}_{\Pr=1} \Pr_L(\mathbf{Y} = A \cup B)}{\Pr_L(A \subseteq \mathbf{Y})} \\
&= \frac{\Pr_L(\mathbf{Y} = A \cup B)}{\Pr_L(A \subseteq \mathbf{Y})} \\
&= \frac{\det(L_{A \cup B})}{\det(L + I_{\bar{A}})}
\end{aligned} \tag{53}$$

6 Sampling DPP:

6.1 express in terms of mixture of elementary DPPs

$$\Pr_L = \frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \mathcal{P}^{V_J} \prod_{n \in J} \lambda_n \quad (54)$$

where $\mathbf{W}_J \equiv \mathbf{W}_{V_J}$ is the associated marginal kernel for \mathcal{P}^{V_J} - we choose to use \mathbf{W}_J instead of K^V , as K is reserved for marginal kernel.

V_J is a set of **orthonormal** vectors, associated with an elementary DPP with marginal kernel $\mathbf{W}_J = \sum_{\mathbf{v} \in V} \mathbf{v} \mathbf{v}^\top$ where $\mathbf{v}_i \in V$ are eigen-vector of L .

6.1.1 advantage of elementary DPP

the most important factor (during first loop) we decides $|J| = |V|$ from by its mixture weight. Then, if we can prove to sample an elementary DPP with marginal kernel \mathbf{W}_J :

$$\Pr_{\mathbf{W}_J}(|\mathbf{Y}| = |J|) = 1 \quad (55)$$

we only need to sample elements of $\{Y_i\}_{i=1}^{|J|}$.

6.1.2 proof for $\Pr_{\mathbf{W}_J}(|\mathbf{Y}| = |J|) = 1$

To begin the proof, we simplify the notation by letting:

$$\mathbf{W}_{V_J} \equiv \mathbf{W}_J \quad (56)$$

Firstly, we know that $\Pr_{\mathbf{W}_J}[|\mathbf{Y}|] = 0 \quad \forall |J| < |\mathbf{Y}|$. Since matrix indexed by \mathbf{Y} will have determinant being zero. However, after we prove that $\mathbb{E}_{\mathbf{W}_J}[|\mathbf{Y}|] = |J|$, so the only way for both to be true is that $|\mathbf{Y}| = |J|$ almost surely:

$$\begin{aligned} \mathbb{E}_{\mathbf{W}_J}[|\mathbf{Y}|] &= \sum_{i=1}^N \mathbb{E}_{\mathbf{W}_J}[\mathbb{1}_{y_i \in \mathbf{Y}}] \\ &= \sum_{i=1}^N \Pr_{\mathbf{W}_J}(y_i \in \mathbf{Y}) \\ &= \sum_{i=1}^N \mathbf{W}_{J,i,i} \quad \text{definition of DPP} \\ &= \text{Tr}(\mathbf{W}_J) \\ &= |J| \end{aligned} \quad (57)$$

Of course, we also need sampling an elementary DPP with **det** (\mathbf{W}_J) kernel has a lot faster computation.

6.2 mixture weight $\frac{\prod_{n \in J} \lambda_n}{\det(L+I)}$

When mixture weights expressed as $\frac{\prod_{n \in J} \lambda_n}{\det(L+I)}$, for example when $J = \{1, 3, 5\}$, its corresponding mixture weights is

$$\frac{\lambda_1 \lambda_3 \lambda_5}{\prod_{n=1}^N (\lambda_n + 1)} \quad (58)$$

6.2.1 probability of including a single element i

knowing the mixture weight of a particular set J isn't that useful, as there are 2^N possible J , i.e., mixture weights.

Operationally, we are more interested in the probability of including the i^{th} \mathbf{v} , i.e., \mathbf{v}_i , by sampling from the **mixture weight** (don't be confused with including $i \in \mathbf{Y}!$). It turns out that:

$$\Pr(\mathbf{v}_i \in V_J) = \frac{\lambda_i}{(\lambda_i + 1)} \quad (59)$$

we demonstrate through an example, let $N = 3$, and we need to decide the inclusion of the element 1:

$$\begin{aligned} \Pr(\mathbf{v}_1 \in V_J) &= \frac{\lambda_1 + \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_1 \lambda_2 \lambda_3}{(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)} \\ &= \frac{\lambda_1(1 + \lambda_2 + \lambda_3 + \lambda_2 \lambda_3)}{(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)} \\ &= \frac{\lambda_1(1 + \lambda_2)(1 + \lambda_3)}{(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)} \\ &= \frac{\lambda_1}{(\lambda_1 + 1)} \end{aligned} \quad (60)$$

since they are exchangeable, we can have:

$$\Pr(\mathbf{v}_i \in V_J) = \frac{\lambda_i}{(\lambda_i + 1)} \quad (61)$$

6.3 sampling \mathcal{P}^V

6.3.1 Elementary DPP:

A DPP is called **elementary** if every eigenvalue of its marginal kernel is $\in \{0, 1\}$

$$1. \text{ example 1: } V \equiv \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$\begin{aligned} \mathbf{W}_J &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= 0 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (62)$$

2. **example 2:** $V \in \left\{ \begin{bmatrix} -0.5735 \\ 0.7781 \\ -0.2562 \end{bmatrix}, \begin{bmatrix} -0.3243 \\ 0.0716 \\ 0.9432 \end{bmatrix}, \begin{bmatrix} 0.7523 \\ 0.6240 \\ 0.2113 \end{bmatrix} \right\}$

$$\begin{aligned} \mathbf{W}_J &= \begin{bmatrix} 0.3945 & -0.0557 & -0.4856 \\ -0.0557 & 0.9949 & -0.0447 \\ -0.4856 & -0.0447 & 0.6106 \end{bmatrix} = 1 \times \begin{bmatrix} -0.5735 \\ 0.7781 \\ -0.2562 \end{bmatrix} \begin{bmatrix} -0.5735 & 0.7781 & -0.2562 \end{bmatrix} \\ &\quad + 0 \times \begin{bmatrix} -0.3243 \\ 0.0716 \\ 0.9432 \end{bmatrix} \begin{bmatrix} -0.3243 & 0.0716 & 0.9432 \end{bmatrix} \\ &\quad + 1 \times \begin{bmatrix} 0.7523 \\ 0.6240 \\ 0.2113 \end{bmatrix} \begin{bmatrix} 0.7523 & 0.6240 & 0.2113 \end{bmatrix} \end{aligned} \quad (63)$$

\mathbf{W}_J is a sum of a set of rank one matrix, each constructed from an ortho-normal set. \mathbf{W}_J is still a valid DPP marginal kernel, although a lot of larger sets will have zero probability.

6.3.2 Multi-Linearity

Lemma 2 Let each \mathbf{W}_n to be rank-one matrix, and sum of $\mathbf{W}_J = \sum_{n \in J} \mathbf{W}_n$:
then we have:

$$\det(\mathbf{W}_J) = \sum_{\underbrace{n_1, n_2, \dots, n_k \in J}_{\text{are distinct}}} \det([(\mathbf{W}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]) \quad (64)$$

RHS can be visualized as when we have a set of $|J|$ matrices $\{\mathbf{W}_n\}_{n=1}^{|J|}$, if we take a column from each of the matrices to form a new matrix \mathbf{W} and to compute its determinant, and then, sum over these determinant of all combinations. Then we get the determinant of the sum of $\{\mathbf{W}_n\}_{n=1}^{|J|}$!
note also that $|J| \geq k$

6.3.3 proof of lemma

write out each column explicitly:

$$\begin{aligned} \det(\mathbf{W}_J) &= \det\left(\left[(\mathbf{W}_J)_1, (\mathbf{W}_J)_2, \dots, (\mathbf{W}_J)_k\right]\right) \\ &= \det\left(\left[\left(\sum_{n \in J} \mathbf{W}_n\right)_1, (\mathbf{W}_J)_2, \dots, (\mathbf{W}_J)_k\right]\right) \quad \text{expand first term} \end{aligned} \quad (65)$$

for example:

$$\begin{aligned} \mathbf{W}_1 &= \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 6 \\ 6 & 4 \end{bmatrix} \\ \mathbf{W}_2 &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \\ \mathbf{W}_J &= \mathbf{W}_1 + \mathbf{W}_2 = \begin{bmatrix} 10 & 8 \\ 8 & 8 \end{bmatrix} \\ \left(\sum_{n \in J} \mathbf{W}_n\right)_1 &= \begin{bmatrix} 10 \\ 8 \end{bmatrix} \end{aligned}$$

because **Multi-linearity** states:

$$\det([\mathbf{a}_1 + \mathbf{b}_1, \mathbf{a}_2, \dots, \mathbf{a}_k]) = \det([\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k]) + \det([\mathbf{b}_1, \mathbf{a}_2, \dots, \mathbf{a}_k]) \quad (66)$$

Therefore,

$$\begin{aligned} \det(\mathbf{W}_J) &= \det\left(\left[\left(\sum_{n \in J} \mathbf{w}_n\right)_1, (\mathbf{W}_J)_2, \dots, (\mathbf{W}_J)_k\right]\right) \\ &= \sum_{n \in J} \det([\mathbf{w}_n)_1, (\mathbf{W}_J)_2, \dots, (\mathbf{W}_J)_k] \end{aligned} \quad (67)$$

Now, we repeat the same thing for the second term and all subsequent terms, But we can't use the same index n for different columns. Therefore, we give a different index $n_i \in J \quad \forall i$:

$$\det(\mathbf{W}_J) = \sum_{n_1 \in J} \sum_{n_2 \in J} \cdots \sum_{n_k \in J} \det(\underbrace{[(\mathbf{W}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]}_{\mathbf{W}}) \quad (68)$$

6.3.4 loop index n_1, \dots, n_k need to be distinct

when we look at:

$$\det(\mathbf{W}_J) = \sum_{n_1 \in J} \sum_{n_2 \in J} \cdots \sum_{n_k \in J} \det([\mathbf{w}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]) \quad (69)$$

not every term is non-zero.

Since \mathbf{W}_n is rank one matrix, $(\mathbf{W}_n)_i$ and $(\mathbf{W}_n)_j$ are linearly dependant. Therefore, the determinant of any matrix containing two or more columns of the **same** \mathbf{W}_n is zero, for example:

$$\det(\mathbf{W}_J) = \det([\mathbf{w}_{n_1})_1, (\mathbf{W}_{n_1})_2, \dots, (\mathbf{W}_{n_k})_k]) = 0 \quad (70)$$

Thus the terms in the sum vanish unless n_1, n_2, \dots, n_k are distinct.

$$\begin{aligned} \det(\mathbf{W}_J) &= \sum_{n_1 \in J} \sum_{n_2 \in J} \cdots \sum_{n_k \in J} \det([\mathbf{w}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]) \\ &= \sum_{\substack{n_1 \in J \\ n_2 \in J \\ \vdots \\ n_k \in J \\ n_1, n_2, \dots, n_k \text{ are distinct}}} \det([\mathbf{w}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]) \\ &= \sum_{\substack{n_1, n_2, \dots, n_k \in J \\ \text{distinct}}} \det([\mathbf{w}_{n_1})_1, (\mathbf{W}_{n_2})_2, \dots, (\mathbf{W}_{n_k})_k]) \end{aligned} \quad (71)$$

6.4 Why mixture of elementary DPPs works

Most importantly, we need to show a DPP with L-ensemble kernel $L = \sum_{n=1}^N \lambda_n v_n v_n^\top$ is a mixture of elementary DPPs:

$$\frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \mathcal{P}^{V_J} \prod_{n \in J} \lambda_n \quad (72)$$

where each \mathcal{P}^{V_J} associate with its own kernel \mathbf{W}_J .

6.4.1 show $\Pr(A \in \mathbf{Y})$ from mixture model also equal $\det(K_A)$

for a particular index set A , we have $k = |A|$ and the associated $\mathbf{W}_n^A = [\mathbf{v}_n \mathbf{v}_n^\top]_A$. This means each of the rank-one matrix of $\mathbf{v}_n \mathbf{v}_n^\top$ gets “chop-off” by the index set A to become \mathbf{W}_n^A . Therefore, we need to show that summation of J (from all the mixture weights) of $\det(\mathbf{W}_J^A)$ gives the right marginal probability $\Pr(A \in \mathbf{Y}) = \det(K_A)$

Start from from mixture of elementary DPPs definition:

$$\begin{aligned}
 \Pr(A \in \mathbf{Y}) &= \frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \det(\mathbf{W}_J^A) \prod_{n \in J} \lambda_n \\
 &= \frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \det\left(\sum_{n \in J} \mathbf{W}_n^A\right) \prod_{n \in J} \lambda_n \quad \text{let } \mathbf{W}_J^A \equiv \mathbf{W}^J \\
 &= \frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \sum_{\substack{n_1, n_2, \dots, n_k \in J \\ \text{distinct}}} \det\left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k\right]\right) \prod_{n \in J} \lambda_n \quad \text{from lemma (2)}
 \end{aligned} \tag{73}$$

For the outer loop, $\sum_{J \subseteq \{1, 2, \dots, N\}}$ when $|J| < k$, then, the inner loop becomes zero. Since it's impossible for $|J| < k$ points to be distinct. Therefore, we need only a subset of $\{1, \dots, N\}$:

$$J \supseteq \{n_1, n_2, \dots, n_k\} \tag{74}$$

By swapping the inner and outer loops, we have:

$$\begin{aligned}
 &= \frac{1}{\det(L + I)} \sum_{J \subseteq \{1, 2, \dots, N\}} \sum_{\substack{n_1, n_2, \dots, n_k \in J \\ \text{distinct}}} \det\left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k\right]\right) \prod_{n \in J} \lambda_n \\
 &= \frac{1}{\det(L + I)} \sum_{\substack{n_1, n_2, \dots, n_k \\ \text{distinct}}} \det\left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k\right]\right) \sum_{J \supseteq \{n_1, n_2, \dots, n_k\}} \prod_{n \in J} \lambda_n
 \end{aligned} \tag{75}$$

For example, let $J \subseteq \{1, 2, 3, 4, 5\}$, and let $\{n_1, n_2, \dots, n_k\} = \{1, 2, 3\}$. Then, $J \supseteq \{n_1, n_2, \dots, n_k\} = \{\{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 4, 5\}\}$:

$$\begin{aligned}
 \sum_{J \supseteq \{n_1, n_2, \dots, n_k\}} \prod_{n \in J} \lambda_n &= \lambda_1 \lambda_2 \lambda_3 + \lambda_1 \lambda_2 \lambda_3 \lambda_4 + \lambda_1 \lambda_2 \lambda_3 \lambda_5 + \lambda_1 \lambda_2 \lambda_3 \lambda_4 \lambda_5 \quad \text{using the example} \\
 &= \lambda_1 \lambda_2 \lambda_3 (1 + \lambda_4 + \lambda_5 + \lambda_4 \lambda_5) \\
 &= \lambda_1 \lambda_2 \lambda_3 (1 + \lambda_4)(1 + \lambda_5) \quad \text{this step is the key} \\
 &= \lambda_1 \lambda_2 \lambda_3 (1 + \lambda_4)(1 + \lambda_5) \frac{(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)(\lambda_4 + 1)(\lambda_5 + 1)}{(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)(\lambda_4 + 1)(\lambda_5 + 1)} \\
 &= \frac{\lambda_1}{\lambda_1 + 1} \frac{\lambda_2}{\lambda_2 + 1} \frac{\lambda_3}{\lambda_3 + 1} (\lambda_1 + 1)(\lambda_2 + 1)(\lambda_3 + 1)(\lambda_4 + 1)(\lambda_5 + 1) \\
 &= \frac{\lambda_{n_1}}{\lambda_{n_1} + 1} \cdots \frac{\lambda_{n_k}}{\lambda_{n_k} + 1} \prod_{n=1}^N (\lambda_n + 1) \quad \text{we generalise it to } N \text{ terms}
 \end{aligned} \tag{76}$$

substituting the expression for $\sum_{J \supseteq \{n_1, n_2, \dots, n_k\}} \prod_{n \in J} \lambda_n$:

$$\begin{aligned}
\Pr_L &= \frac{1}{\det(L+I)} \sum_{\substack{n_1, n_2, \dots, n_k \\ \text{distinct}}}^N \det \left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k \right] \right) \sum_{J \supseteq \{n_1, n_2, \dots, n_k\}} \prod_{n \in J} \lambda_n \\
&= \frac{1}{\prod_{n=1}^N (\lambda_n + 1)} \sum_{\substack{n_1, n_2, \dots, n_k \\ \text{distinct}}}^N \det \left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k \right] \right) \frac{\lambda_{n_1}}{\lambda_{n_1} + 1} \cdots \frac{\lambda_{n_k}}{\lambda_{n_k} + 1} \prod_{n=1}^N (\lambda_n + 1) \\
&= \sum_{\substack{n_1, n_2, \dots, n_k \\ \text{distinct}}}^N \det \left(\left[(\mathbf{W}_{n_1}^A)_1, (\mathbf{W}_{n_2}^A)_2, \dots, (\mathbf{W}_{n_k}^A)_k \right] \right) \frac{\lambda_{n_1}}{\lambda_{n_1} + 1} \cdots \frac{\lambda_{n_k}}{\lambda_{n_k} + 1} \\
&= \sum_{\substack{n_1, n_2, \dots, n_k \\ \text{distinct}}}^N \det \left(\left[(\mathbf{W}_{n_1}^A)_1 \frac{\lambda_{n_1}}{\lambda_{n_1} + 1}, (\mathbf{W}_{n_2}^A)_2 \frac{\lambda_{n_2}}{\lambda_{n_2} + 1}, \dots, (\mathbf{W}_{n_k}^A)_k \frac{\lambda_{n_k}}{\lambda_{n_k} + 1} \right] \right) \\
&\quad \because \alpha \beta \det([\mathbf{a}_1 \quad \mathbf{a}_2]) = \det([\alpha \mathbf{a}_1 \quad \beta \mathbf{a}_2]) \\
&\quad \because \text{every column of } \mathbf{W}_n \text{ times by } \frac{\lambda_n}{\lambda_n + 1} \implies \text{whole } \mathbf{W}_n \text{ times by } \frac{\lambda_n}{\lambda_n + 1} \\
&= \det \left(\sum_{n=1}^N \frac{\lambda_n}{\lambda_n + 1} \mathbf{W}_n^A \right) \quad \text{apply lemma (2) again, with } J \equiv \{1, \dots, N\} \\
&= \det(K_A) \quad \text{using Eq.(41) by noting } \sum_{n=1}^N \frac{\lambda_n}{\lambda_n + 1} \mathbf{W}_n = K
\end{aligned} \tag{77}$$