A survey on Strongly Rayleigh measures and their mixing time analysis

Raghav Somani & Mitchell Wortsman

January 2, 2020

1 Negative Association and Negative dependence

Negative association informally refers to the notion of *diversity* when we are interested in sampling a subset of elements from the universe set. Before describing this in detail, let us define the notion of positive association

Definition 1.1. A probability measure μ is said to be *positively associated* if for every pair of increasing real valued set functions f and g on $[n]$, we have

$$\int fgd\mu \geq \int fd\mu \cdot \int gd\mu \tag{1.1}$$

The definition of positive association is implied by a more general *positive lattice condition*, also known as the FKG lattice condition [5]

$$\mu(S)\mu(T) \leq \mu(S \cup T)\mu(S \cap T) \tag{1.2}$$

If we sample $S \in [n]$ according to μ and denote $X_i = \mathbb{1}[i \in S]$, then from Equation (1.1) we get

$$\mathbb{E}[X_i X_j] \geq \mathbb{E}[X_i]\,\mathbb{E}[X_j] \qquad \forall\, i,j \in [n]\,. \tag{1.3}$$

The usual definition of *negative association* [7] in contrast, is not robust and does not follow similar implications like (1.2) $\implies$ (1.1).

Definition 1.2. A probability measure μ is said to be *negatively associated* if for every pair of increasing real valued set functions f and g on $[n]$, provided that f and g depend on disjoint subsets, then we have

$$\int fgd\mu \geq \int fd\mu \cdot \int gd\mu \tag{1.4}$$

For negative dependence, we do not have similar implications like Equation (1.3) because of the identity

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i,j=1}^{n} \left(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]\right) \tag{1.5}$$

which is always non-negative. For positive association, we have each term in the summation in Equation (1.5) non-negative. However, for negative association not all summands can be simultaneously non-negative if the variance of the sum is strictly positive. This is precisely taken care by the constraints put on f and g in Definition 1.2 as since f is always positively correlated to f.

There are several qualitative notions of *negative dependence* studied in the literature, but negative association has several distinct advantages when compared to others. We however would like to have some ideal properties to hold in our notion of negative dependence in addition to negative association, like closedness under natural operations like marginalization, product, projection, relabeling, and properties which extend to negative correlation and external field [9]. Apart from these, we might also want to have properties like stochastic coverage, ultra log-concavity of rank sequence $P_{S\sim\mu}\{|S| = k\}$. These properties were studied in [2] using the concept of generating polynomials which give rise to the family of Strongly-Rayleigh measures which satisfy all the desired properties of negative dependence which we would like to have.

2 Strongly Rayleigh probability measures

The class of *Strongly Rayleigh measures*, introduced in [2] as we will see, can characterize all the desired properties which we would like to have from the notion of negative dependence as discussed in the previous section. This comes from some geometric conditions on the generating polynomial of a measure.

2.1 Generating polynomials

For $i \in [n]$, consider the i-th coordinate function on $2^{[n]}$ to be the binary random variable given by $X_i(S) = \mathbb{1}[i \in S]$, and the characteristic function $\chi_S(T) = \mathbb{1}[S = T]$ where $S, T \in [n]$. Therefore any scalar function $f : 2^{[n]} \to \mathbb{R}$ can be written as $f = \sum_{S \in 2^{[n]}} f(S)\chi_S$, and can be viewed as a multi-affine polynomial $f(X_1, X_2, \ldots, X_n)$ of the random variables $\{X_i\}_{i=1}^n$ satisfying $f(\mathbf{1}) = 1$. If μ is a distribution on $2^{[n]}$, then its generating polynomial is defined as

$$g_\mu(z) = \int z^S d\mu(S) = \sum_{S \in 2^{[n]}} \mu(S) z^S, \quad z = (z_i)_{i=1}^n, \quad z^S := \prod_{i \in S} z_i, \quad z_i \in \mathbb{C} \ \forall \ i \in [n]. \tag{2.1}$$

Therefore for any given function $f : 2^{[n]} \to \mathbb{C}$ of the form $f = \sum_{S \in [n]} a_S z^S$ with $a_{[n]} = 1$ and $a_S \in [0, 1] \ \forall \ S \in [n]$, its corresponding measure μ_f on $2^{[n]}$ can be defined by setting $\mu_f(S) = a_S \ \forall \ S \in [n]$. This gives us a one-to-one correspondence between $\mathcal{P}_n$, the set of measures on $2^{[n]}$, and $\mathfrak{P}_n$, the set of multi affine polynomials in n variables $f(z_1, z_2, \ldots, z_n)$ with non-negative coefficients such that $f(\mathbf{1}) = 1$. Precisely,

$$g_{\mu_f} = f \quad \text{and} \quad \mu_{g_\nu} = \nu \qquad \forall \ \nu \in \mathfrak{P}_n \text{ and } f \in \mathcal{P}_n. \tag{2.2}$$

This one-to-one mapping between $\mathfrak{P}_n$ and $\mathcal{P}_n$ via generating polynomial gives us a strong tool to algebraically and geometrically analyze the properties on negative dependence. Generating polynomials can be shown to be closed under product, projection, conditioning, symmetrization, imposition of external field, partial symmetrization and truncation.

2.2 Strongly Rayleigh measures

Definition 2.1. A measure $\mu \in \mathfrak{P}_n$ is said to be *pairwise negatively correlated* if $\forall \ 1 \leq i \neq j \leq n$,

$$\text{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] \leq 0 \tag{2.3}$$

which in the language of generating polynomials implies

$$\frac{\partial g_\mu}{\partial z_i}(\mathbf{1}) \frac{\partial g_\mu}{\partial z_j}(\mathbf{1}) \geq \frac{\partial^2 g_\mu}{\partial z_i \partial z_j}(\mathbf{1}) g_\mu(\mathbf{1}). \tag{2.4}$$

Rayleigh polynomials have a stronger notion of pairwise negative correlation and satisfy numerous closedness properties [2].

Definition 2.2. A polynomial $f \in \mathcal{P}_n$ is called a *Rayleigh polynomial* if

$$\frac{\partial f}{\partial z_i}(x) \frac{\partial f}{\partial z_j}(x) \geq \frac{\partial^2 f}{\partial z_i \partial z_j}(x) f(x) \quad \forall \ x \in \mathbb{R}_+^n \text{ and } \ \forall \ i, j \in [n]. \tag{2.5}$$

Definition 2.3. A measure $\mu \in \mathfrak{P}_n$ is said to be a *Rayleigh measure* if its generating polynomial g_μ is a Rayleigh polynomial.

Definition 2.4. A polynomial $f \in \mathbb{C}[z_1, z_2, \ldots, z_n]$ is called *real stable* is all its coefficients are real and $\Im(z_i) > 0 \ \forall \ i \in [n] \implies g_\mu(z) \neq 0$.

Definition 2.5. A measure $\mu \in \mathfrak{P}_n$ is *strongly Rayleigh* if its generating polynomial g_μ is real stable. A real stable polynomial $f \in \mathcal{P}_n$ is called a *strongly Rayleigh polynomial*.

It can be shown that a strongly Rayleigh measure is necessarily a Rayleigh measure [3]. A polynomial f is real stable if and only if $\forall \ a \in \mathbb{R}_+^n$, and $b \in \mathbb{R}^n$, the uni-variate polynomial $f(a_1 t + b_1, \ldots, a_n t + b_n)$ is real rooted. This can be used to check if a polynomial is strongly Rayleigh. Real stable polynomials satisfy some interesting closure properties. Let $\mathcal{H}_n(\mathbb{R})$ be the set of all real stable polynomials in n variables, and let $f \in \mathcal{H}_n(\mathbb{R})$ have degree d_j in variable $z_j \ \forall \ j \in [n]$, then

1. $\partial_j f \in \mathcal{H}_n(\mathbb{R}) \cap \{0\} \ \forall\ j \in [n]$,
2. $f(z_1, \ldots, z_{j-1}, \alpha z_j, z_{j+1}, \ldots, z_n) \in \mathcal{H}_n(\mathbb{R})\ \forall\ j \in [n]$ and $\alpha > 0$,
3. $f(z_1, \ldots, z_{j-1}, \beta, z_{j+1}, \ldots, z_n) \in \mathcal{H}_{n-1}(\mathbb{R}) \cap \{0\}\ \forall\ j \in [n]$ and $\beta \in \mathbb{R}$,
4. $\prod_{j=1}^{n} z_j^{d_j} \cdot f(\lambda_1 z_1^{-1}, \ldots, \lambda_n z_n^{-1}) \in \mathcal{H}_n(\mathbb{R})$ if $\pm(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}_+^n$,
5. $f(z_1, \ldots, z_{i-1}, z_j, z_{i+1}, \ldots, z_n) \in \mathcal{H}_{n-1}(\mathbb{R})\ \forall\ i \neq j \in [n]$, and
6. If $\{f_j\}_{j=1}^{\infty} \subset \mathcal{H}_n(\mathbb{R})$ and $f \in \mathbb{R}[z_1, \ldots, z_n] \setminus \{0\}$ is the limit, uniformly on compact subsets of $\mathbb{C}^n$, of the sequence $\{f_j\}_{j=1}^{\infty}$, then $f \in \mathcal{H}_n(\mathbb{R})$.

We will now slowly turn up to some important classes of strongly Rayleigh measures which are nicer to work wit, both algebraically and computationally.

Proposition 2.6. *If $\{\mathbf{A}_i\}_{i=1}^m$ and $\mathbf{B}$ are (complex) positive semi-definite (PSD) matrices in $\mathbb{C}^{n\times n}$, then the polynomial*

$$\det\left(\sum_{i=1}^{m} z_i \mathbf{A}_i + \mathbf{B}\right)$$

is either identically zero or real stable with all non-negative coefficients.

Proof. We can use the test to check if a polynomial is strongly Rayleigh by defining $\mathbf{z}(t) = \boldsymbol{\lambda} t + \boldsymbol{\mu}$ with $\boldsymbol{\lambda} \in \mathbb{R}_+^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $t \in \mathbb{C}$. Plugging this in, we have

$$\det\left(\sum_{i=1}^{m} z_i \mathbf{A}_i + \mathbf{B}\right) = \det\left(\left(\sum_{i=1}^{m} \lambda_i \mathbf{A}_i\right) t + \left(\sum_{i=1}^{m} \mu_i \mathbf{A}_i + \mathbf{B}\right)\right) \tag{2.6}$$

Defining $\mathbf{P} := \sum_{i=1}^{m} \lambda_i \mathbf{A}_i$, and $\mathbf{H} := \sum_{i=1}^{m} \mu_i \mathbf{A}_i + \mathbf{B}$, we have that $\mathbf{P}$ is positive definite and therefore is invertible and has a square root, lets say $\mathbf{Q}$. Therefore,

$$\det\left(\sum_{i=1}^{m} z_i \mathbf{A}_i + \mathbf{B}\right) = \det(\mathbf{P}) \det(t\mathbf{I} + \mathbf{Q}\mathbf{H}\mathbf{Q}^*) \tag{2.7}$$

is a positive constant multiple of a characterstic polynomial in t and thus has all real zeros and non-negative coefficients. □

Using this, we directly have that $\det(\mathbf{A} + \mathbf{Z})$ is a multi-affine real stable polynomial with all non-negative coefficients for any PSD matrix $\mathbf{A}$ and $\mathbf{Z} := \operatorname{diag}(z_1, \ldots, z_n)$.
Given $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $S \in [n]$, define $\mathbf{A}[S]$ the principle sub-matrix with rows and columns indexed by S. Let $\mathbf{A}\langle S\rangle := \det\left(\mathbf{A}\left[\bar{S}\right]\right)$ be the principal minor of $\mathbf{A}$ with rows and columns indexed by $\bar{S}$, with $\mathbf{A}\langle[n]\rangle := 1$. A matrix A is called a P-matrix if all its subminors are positive.

Definition 2.7 (Hadamard-Fischer Inequalities and GKK-matrix)**.** A P-matrix $\mathbf{A}$ is a GKK-matrix if it satisfies the Hadamard-Fischer Inequality

$$\mathbf{A}\langle S\rangle\, \mathbf{A}\langle T\rangle \geq \mathbf{A}\langle S \cup T\rangle\, \mathbf{A}\langle S \cap T\rangle \tag{2.8}$$

or equivalently, a P-matrix $\mathbf{A}$ is GKK if and only if the probability measure defined by $\mu_{\mathbf{A}}$ on $2^{[n]}$ defined by

$$\mu_{\mathbf{A}}(S) = \frac{\mathbf{A}\langle S\rangle}{\sum_{T\subseteq[n]} \mathbf{A}\langle T\rangle} \tag{2.9}$$

satisfies the negative lattice condition. The probability measure $\mu_{\mathbf{A}}$ however, has a closed form which can be computed in polynomial time using the below theorem

Theorem 2.8. *For any $U \subseteq [n]$,*

$$\sum_{U \subseteq T \subseteq \mathcal{X}} \det\left(\mathbf{A}\left[T\right]\right) = \det\left(\mathbf{A} + \mathbf{I}_{\bar{U}}\right) \tag{2.10}$$

where $\left(\mathbf{I}_{\bar{U}}\right)_{i,i} = \mathbb{1}\left[i \in \bar{U}\right]\ \forall\ i \in \mathcal{X}$, and zeros everywhere else.

Proof. The proof follows and inductive argument. Suppose $U = \mathcal{X}$, then the statement trivially holds. Now suppose it holds whenever $\left|\bar{U}\right| < k$. Given U such that $\left|\bar{U}\right| = k > 0$, let i be an element of $\mathcal{X}$ where $i \in \bar{U}$. We can write $\mathbf{A} + \mathbf{I}_{\bar{U}}$ as

$$\mathbf{A} + \mathbf{I}_{\bar{U}} = \begin{bmatrix} A_{ii} + 1 & \mathbf{A}_{i\bar{i}} \\ \mathbf{A}_{\bar{i}i} & \mathbf{A}_{\mathcal{X}\setminus\{i\}} + \mathbf{I}_{(\mathcal{X}\setminus\{i\})\setminus U} \end{bmatrix} \tag{2.11}$$

where $\mathbf{A}_{\bar{i}i}$ is the sub-column of the i-th column of $\mathbf{A}$ whose rows correspond to $\bar{i}$ and similarly $\mathbf{A}_{i\bar{i}}$. From the multi-linearity property of determinant, we have

$$\det\left(\mathbf{A} + \mathbf{I}_{\bar{U}}\right) = \begin{vmatrix} A_{ii} + 1 & \mathbf{A}_{i\bar{i}} \\ \mathbf{A}_{\bar{i}i} & \mathbf{A}_{\mathcal{X}\setminus\{i\}} + \mathbf{I}_{(\mathcal{X}\setminus\{i\})\setminus U} \end{vmatrix} \tag{2.12}$$

$$= \begin{vmatrix} A_{ii} & \mathbf{A}_{i\bar{i}} \\ \mathbf{A}_{\bar{i}i} & \mathbf{A}_{\mathcal{X}\setminus\{i\}} + \mathbf{I}_{(\mathcal{X}\setminus\{i\})\setminus U} \end{vmatrix} + \begin{vmatrix} 1 & \mathbf{0} \\ \mathbf{A}_{\bar{i}i} & \mathbf{A}_{\mathcal{X}\setminus\{i\}} + \mathbf{I}_{(\mathcal{X}\setminus\{i\})\setminus U} \end{vmatrix} \tag{2.13}$$

$$= \det\left(\mathbf{A} + \mathbf{I}_{\overline{U\cup\{i\}}}\right) + \det\left(A_{\mathcal{X}\setminus\{i\}} + \mathbf{I}_{(\mathcal{X}\setminus\{i\})\setminus U}\right) \tag{2.14}$$

From the induction assumption, we get

$$\det\left(\mathbf{A} + \mathbf{I}_{\bar{U}}\right) = \sum_{U\cup\{i\} \subseteq T \subseteq \mathcal{X}} \det\left(\mathbf{A}\left[T\right]\right) + \sum_{U \subseteq T \subseteq \mathcal{X}\setminus\{i\}} \det\left(\mathbf{A}\left[T\right]\right) \tag{2.15}$$

$$= \sum_{U \subseteq T \subseteq \mathcal{X}} \det\left(\mathbf{A}\left[T\right]\right) \tag{2.16}$$

where the sum can be written as all the sets which include i and all the sets which do not, completing the proof. □

For $U = \emptyset$, we get a compact form for the measure $\mu_{\mathbf{A}}$,

$$\mu_{\mathbf{A}}(S) = \frac{\mathbf{A}\left\langle S\right\rangle}{\det\left(\mathbf{A} + \mathbf{I}\right)} \tag{2.17}$$

$$= \frac{\det\left(\mathbf{A}\left[\bar{S}\right]\right)}{\det\left(\mathbf{A} + \mathbf{I}\right)} \tag{2.18}$$

Therefore the generating polynomial of the measure $\mu_{\mathbf{A}}$ becomes

$$g_{\mu_{\mathbf{A}}}(z) = \det\left(\mathbf{A} + \mathbf{I}\right)^{-1} \sum_{S \subseteq [n]} \mathbf{A}\left\langle S\right\rangle z^S = \det\left(\mathbf{A} + \mathbf{I}\right)^{-1} \det\left(\mathbf{A} + \mathbf{Z}\right), \tag{2.19}$$

where $\mathbf{Z}$ is a diagonal matrix with $Z_{ii} = z_i\ \forall\ i \in [n]$ and $z^S = \prod_{i \in S} z_i$.

Definition 2.9. A P-matrix $\mathbf{A}$ is Rayleigh its associated probability matrix $\mu_{\mathbf{A}}$ is Rayleigh.

Theorem 2.10. $\mathbf{A}$ *is a Rayleigh matrix* $\iff \mathbf{A} + \mathbf{Z}$ *is a GKK matrix for all positive diagonal matrices* $\mathbf{Z}$.

Proof. $\Longleftarrow$) $\mathbf{A} + \mathbf{Z}$ is GKK

$$\implies \left(\mathbf{A} + \mathbf{Z}\right)\left\langle\{i\}\right\rangle \cdot \left(\mathbf{A} + \mathbf{Z}\right)\left\langle\{j\}\right\rangle \geq \left(\mathbf{A} + \mathbf{Z}\right)\left\langle\{i,j\}\right\rangle \cdot \left(\mathbf{A} + \mathbf{Z}\right)\left\langle\emptyset\right\rangle \qquad \forall\ i,j \in [n] \tag{2.20}$$

$$\implies \frac{\partial \det\left(\mathbf{A} + \mathbf{Z}\right)}{\partial z_i} \frac{\partial \det\left(\mathbf{A} + \mathbf{Z}\right)}{\partial z_j} \geq \frac{\partial^2 \det\left(\mathbf{A} + \mathbf{Z}\right)}{\partial z_i \partial z_j} \cdot \det\left(\mathbf{A} + \mathbf{Z}\right) \qquad \forall\ i,j \in [n] \tag{2.21}$$

$$\implies \frac{\partial g_{\mu_{\mathbf{A}}}(z)}{\partial z_i} \frac{\partial g_{\mu_{\mathbf{A}}}(z)}{\partial z_j} \geq \frac{\partial^2 g_{\mu_{\mathbf{A}}}(z)}{\partial z_i \partial z_j} \cdot g_{\mu_{\mathbf{A}}}(z) \qquad \forall\ i,j \in [n] \tag{2.22}$$

$\implies \mu_{\mathbf{A}}$ is Rayleigh $\implies \mathbf{A}$ is a Rayleigh matrix.

$\implies$) If $\mathbf{A}$ is Rayleigh, then $\mathbf{A}[S]$ is also Rayleigh for $U \subseteq [n]$. Let $|U| \geq 2$, and $U := S \cup \{i,j\}$ where $U \cap \{i,j\} = \emptyset$ from which we get

$$\det(\mathbf{A}+\mathbf{Z})[U \setminus \{j\}] \det(\mathbf{A}+\mathbf{Z})[U \setminus \{i\}] \geq \det(\mathbf{A}+\mathbf{Z})[U \setminus \{i,j\}] \det(\mathbf{A}+\mathbf{Z})[U] \tag{2.23}$$

$$\implies \det(\mathbf{A}+\mathbf{Z})[S \cup \{i\}] \det(\mathbf{A}+\mathbf{Z})[S \cup \{j\}] \geq \det(\mathbf{A}+\mathbf{Z})[S] \det(\mathbf{A}+\mathbf{Z})[S \cup \{i,j\}] \tag{2.24}$$

Let $k, l, \{S_i\}_{i=0}^{k}$ and $\{T_i\}_{j=0}^{l}$ be such that

$$S \cap T := S_0 \subset S_1 \subset \ldots \subset S_k =: S \tag{2.25}$$

$$S \cap T := T_0 \subset T_1 \subset \ldots \subset T_l =: T \tag{2.26}$$

which gives us that for all $(i,j) \in [k] \times [l]$,

$$\det(\mathbf{A}+\mathbf{Z})[S_i \cup T_{j-1}] \det(\mathbf{A}+\mathbf{Z})[S_{i-1} \cup T_j] \geq \det(\mathbf{A}+\mathbf{Z})[S_{i-1} \cup T_{j-1}] \det(\mathbf{A}+\mathbf{Z})[S_i \cup T_j] \tag{2.27}$$

Taking product over all such equations (2.27) we get

$$\det(\mathbf{A}+\mathbf{Z})[S] \det(\mathbf{A}+\mathbf{Z})[T] \geq \det(\mathbf{A}+\mathbf{Z})[S \cup T] \det(\mathbf{A}+\mathbf{Z})[S \cap T] \qquad \forall\ S, T \subseteq [n] \tag{2.28}$$

$$\implies (\mathbf{A}+\mathbf{Z})\langle S \rangle (\mathbf{A}+\mathbf{Z})\langle T \rangle \geq (\mathbf{A}+\mathbf{Z})\langle S \cup T \rangle (\mathbf{A}+\mathbf{Z})\langle S \cap T \rangle \tag{2.29}$$

$\implies$ $\mathbf{A}+\mathbf{Z}$ is a GKK matrix.

□

3 Determinantal Probability measures

A widely used notion of negative dependence is via a class of measures called determinantal probability measures which is closely associated with negative association and strongly Rayleighness [8].

Definition 3.1 (Determinantal probability measure)**.** A measure $\mu \in \mathfrak{P}_n$ is a determinantal probability measure if there is a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ such that for any subset $S \subseteq [n]$, we have

$$\mu(\{T : S \subseteq T\}) = \det(\mathbf{A}[S]) \tag{3.1}$$

Definition 3.2 (Positive contraction)**.** A positive semidefinite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called a *positive contraction* if $\|\mathbf{A}\|_2 \leq 1$, where $\|\cdot\|_2$ is the supremum (operator) norm.

Theorem 3.3. *If μ is a determinantal measure on $2^{[n]}$ whose corresponding matrix is apositive contraction, then μ is strongly Rayleigh.*

Proof. Since positive definite matrices are dense in the space of all all positive semi-definite matrices, it is enough to proove the statement for invertible positive definite matrices. The generating polynomial of a determinantal measure is

$$g_\mu(z) = \det(\mathbf{I} - \mathbf{A} + \mathbf{A}\mathbf{Z}) = \det(\mathbf{A}) \cdot \det\left(\mathbf{A}^{-1} - \mathbf{I} + \mathbf{Z}\right), \tag{3.2}$$

where $\mathbf{Z} = \operatorname{diag}(z_1, \ldots, z_n)$. Since $\mathbf{A}^{-1} - \mathbf{I}$ is positive semi-definite and $\det(A) > 0$, so from 2.6, polynomial $g_\mu(z)$ is real stable therefore μ is strongly Rayleigh. □

Definition 3.4 (Determinantal Point Process (DPP))**.** A discrete *determinantal point process* is a stochastic point process whose probability distribution μ is characterized by a determinant of some positive semi-definite ensemble matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ where for any $S \subseteq [n]$

$$\mathbb{P}[S] \propto \det(\mathbf{L}_S) \tag{3.3}$$

Definition 3.5 (k-DPP)**.** For $k \in [n] \cup \{0\}$, the truncation of a DPP measure μ to k, μ_k is called a k-DPP.

4 Monte Carlo Markov Chain Algorithms for Sampling

We now review Monte Carlo Markov Chain Algorithms (MCMC) for sampling strongly Rayleigh distributions and determinantal point processes [1]. Notably, Anari, Oveis Gharan, and Rezaei provide and prove a $\text{poly}(k)\mathcal{O}\left(n \log\left(n/\epsilon\right)\right)$ algorithm for sampling from a k-DPP where ϵ is the error in total variation distance. We first provide the necessary background on Markov Chains and MCMC, then discuss the use of MCMC methods for sampling Strongly Rayleigh measures and k-DPPs.
We dive into the parts of the paper which we find interesting but are not covered in too much detail by Anari, Oveis Gharan, and Rezaei. We provide more brief descriptions for the proofs covered in detail by their work.

4.1 Markov Chains and MCMC

4.1.1 Preliminaries

A Markov Chain formalizes a process which is quite natural: a process transitioning from state to state in which the current state determines the future dynamics of the system.

Definition 4.1. A Markov Chain is a collection of random variables $\{X_t\}$ for $X_t \in \Omega$, $t \in \mathbb{N}$ which satisfies

$$\mathbb{P}\left(X_t = x_t \mid \{X_i\}_{i<t} = \{x_i\}_{i<t}\right) = \mathbb{P}\left(X_t = x_t \mid X_{t-1} = x_{t-1}\right). \tag{4.1}$$

Equivalently, the future is conditionally independent of the past given the present.

Definition 4.2. A Markov Chain is said to be *time homogeneous* if $\mathbb{P}\left(X_t = x \mid X_{t-1} = y\right)$ does not depend on t for all $x, y \in \Omega$. A *time homogeneous* Markov Chain is defined by a *Markov kernel* P where

$$P(x, y) = \mathbb{P}\left(X_t = x \mid X_{t-1} = y\right) \tag{4.2}$$

Definition 4.3. The *stationary distribution* of a Markov Chain is π if

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y) \tag{4.3}$$

for all $y \in \Omega$.

For the remainder of this work we will only consider time-homogeneous Markov Chains (Ω, P, π), where Ω is the state space, P is the Markov kernel, and π is the stationary distribution.

Definition 4.4. A Markov Chain (Ω, P, π) is said to be reversible if

$$\pi(x)P(x, y) = \pi(y)P(y, x) \tag{4.4}$$

for all $x, y \in \Omega$. Equation 4.4 is referred to in literature as the *detailed balance condition*.

Theorem 4.5. *If a Markov Chain $M = (\Omega, P, \pi)$ satisfies the detailed balance condition with distribution ν, then ν is the stationary distribution of M. Formally, if $\nu(x)P(x, y) = \nu(y)P(y, x)$ for all $x, y \in \Omega$ then $\nu = \pi$*

Proof. Consider any $y \in \Omega$. From the detailed balance condition we may write

$$\sum_{x \in \Omega} \nu(x)P(x, y) = \sum_{x \in \Omega} \nu(y)P(y, x) = \nu(y) \sum_{x \in \Omega} P(y, x) = \nu(y) \tag{4.5}$$

□

Definition 4.6. A Markov Chain (Ω, P, π) is said to be *lazy* if $P(x, x) \geq 1/2$ for all $x \in \Omega$.

4.1.2 Mixing Times

Definition 4.7. The *Total Variation Distance* of distributions $\pi, \nu : \Omega \to \mathbb{R}_+$ is given by

$$\|\nu - \pi\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\nu(x) - \pi(x)|. \tag{4.6}$$

Definition 4.8. For $x \in \Omega$ and $\epsilon > 0$, the *Mixing Time* $\tau_x(\epsilon)$ is the number of steps required for a Markov chain (Ω, P, π) with start state x to be approximately distributed according to π. Formally, if $P^t(x, \cdot)$ is the distribution (at time t) of the chain which starts at x, then

$$\tau_x(\epsilon) = \min\{t \ : \ \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \epsilon\} \tag{4.7}$$

Definition 4.9. The *Poincare Constant* λ of a Markov Chain (Ω, P, π) is given by

$$\lambda = \inf_{f:\Omega \to \mathbb{R}} \frac{\mathcal{E}(f, f)}{\text{Var}_\pi(f)} \tag{4.8}$$

where the *Dirchlet Form* $\mathcal{E}(f, f)$ is defined to be

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x,y \in \Omega} (f(x) - f(y))^2 P(x, y)\pi(x) \tag{4.9}$$

and the *Variance* $\text{Var}_\pi(f)$ is given by

$$\text{Var}_\pi(f) = \sum_{x \in \Omega} (f(x) - \mathbb{E}_\pi f)^2 \pi(x) \tag{4.10}$$

and the infimum is over all functions with nonzero variance.

Lemma 4.10. *A reversible Markov Chain with $\Omega = \{0, 1\}$ and $P(0, 1) = c \cdot \pi(1)$ satisfies $\lambda = c$.*

Proof. The chain is reversible and so

$$P(1, 0) = \frac{P(0, 1) \cdot \pi(0)}{\pi(1)} = c \cdot \pi(0). \tag{4.11}$$

We then have

$$\lambda = \inf_{f:\Omega \to \mathbb{R}} \frac{1}{2} \frac{(f(0) - f(1))^2 (P(0, 1)\pi(0) + P(1, 0)\pi(1))}{(f(0) - \mathbb{E}_\pi f)^2 \pi(0) + (f(1) - \mathbb{E}_\pi f)^2 \pi(1)} \tag{4.12}$$

$$= \inf_{f:\Omega \to \mathbb{R}} c \cdot \pi(0) \cdot \pi(1) \frac{(f(0) - f(1))^2}{(f(0) - \mathbb{E}_\pi f)^2 \pi(0) + (f(1) - \mathbb{E}_\pi f)^2 \pi(1)}. \tag{4.13}$$

Notice that if we add any constant k to f then both the numerator and denominator are unchanged. Without loss of generality we may therefore pick f so that $\mathbb{E}_\pi f = 0$. We then have

$$\lambda = \inf_{f:\Omega \to \mathbb{R}} c \cdot \pi(0) \cdot \pi(1) \frac{(f(0) - f(1))^2}{f(0)^2 \pi(0) + f(1)^2 \pi(1)}. \tag{4.14}$$

Additionally, notice that if we scale f by k then we may factor out k^2 from both the numerator and denominator in the equation above. Without loss of generality we may therefore scale f so that $f(0) = \pi(1)$. Since $\mathbb{E}_\pi f = 0$ we have that $f(1) = -\pi(0)$ and

$$\lambda = c \cdot \pi(0) \cdot \pi(1) \frac{(\pi(0) + \pi(1))^2}{\pi(1)^2 \pi(0) + \pi(0)^2 \pi(1)} = c \cdot \frac{(\pi(0) + \pi(1))^2}{\pi(1) + \pi(0)} = c \tag{4.15}$$

since $\pi(0) + \pi(1) = 1$. □

Lemma 4.11. *The largest eigenvalue of Markov kernel P is 1.*

Proof. Since P is a Markov kernel, $0 \leq P(i,j) \leq 1$ for each i,j and each row sum is 1. We first show that P has eigenvalue 1. Let $\mathbf{1}$ be a vector of ones. Then $P\mathbf{1} = \mathbf{1}$ as row i of $P\mathbf{1}$ is $\sum_j P(i,j) = 1$.
Assume for sake of contradiction that $Px = \lambda x$ for $\lambda > 1$. Let $i = \text{argmax}_j (Px)_j$. □

Lemma 4.12. *λ is an eigenvalue of P with eigenvector v iff $1-\lambda$ is an eigenvalue of P with eigenvector v.*

Proof. If λ is an eigenvalue of P with eigenvecotr v then

$$(I-P)v = v - Pv = v - \lambda v = (1-\lambda)v \tag{4.16}$$

and similarly if $(1-\lambda)$ is an eigenvalue of $I-P$ with eigenvector v then

$$Pv = Pv + v - v = -(I-P)v + v = -(1-\lambda)v + v = \lambda v. \tag{4.17}$$

□

Theorem 4.13. *The second largest eigenvalue of a Markov chain M with kernel P is $1-\lambda$, where λ is the Poincare constant.*

Proof. For brevity we will prove a slightly restricted version of this theorem. Namely, we will assume that $P(x,x) = 1/2$ for all $x \in \Omega$ (we see later that we are roughly concerned with this case). We use the standard inner product for $L^2(\pi)$,

$$\langle f, g \rangle_\pi = \mathbb{E}_\pi [f \cdot g] = \sum_{x \in \Omega} \pi(x) f(x) g(x). \tag{4.18}$$

By Lemma 4.12 may equivalently find the second smallest eigenvalue of $I-P$, which we denote λ_2. By Lemma 4.11 and the Rayleigh Criterion we have

$$\lambda_2 = \inf_{\substack{f:\Omega \to \mathbb{R} \\ \langle f, \mathbf{1} \rangle = 0}} \frac{\langle f, (I-P)f \rangle_\pi}{\langle f, f \rangle_\pi}. \tag{4.19}$$

Expanding the numerator of Equation 4.19 and using $P(x,y)\pi(x) = P(y,x)\pi(y)$ we have

$$\langle f, (I-P)f \rangle_\pi = \sum_{x \in \Omega} \underbrace{(1 - P(x,x))}_{=P(x,x)} f(x)^2 \pi(x) - \sum_{x \neq y} P(x,y) f(x) f(y) \pi(x) \tag{4.20}$$

$$= \sum_{x \in \Omega} P(x,x) f(x)^2 \pi(x) - 2 \sum_{x < y} P(x,y) f(x) f(y) \pi(x) \tag{4.21}$$

$$= \sum_{x,y \in \Omega} P(x,y) \pi(x) (f(x) - f(y))^2 = \mathcal{E}(f,f) \tag{4.22}$$

and likewise the denominator yields

$$\langle f, f \rangle_\pi = \sum_{x \in \Omega} f(x)^2 \pi(x) = \mathbb{E}_\pi \left[f^2 \right] \tag{4.23}$$

which is equal to $\text{Var}_\pi(f)$ since the condition $\langle f, \mathbf{1} \rangle_\pi$ guarantees that $\sum_x f(x)\pi(x) = \mathbb{E}[X] = 0$. □

Theorem 4.14. *For any reversible lazy Markov chain $M = (\Omega, P, \pi)$ with Poincare constant λ, $\epsilon > 0$ and $x \in \Omega$,*

$$\tau_x(\epsilon) \leq \frac{1}{\lambda} \cdot \log \left(\frac{1}{\epsilon \cdot \pi(x)} \right). \tag{4.24}$$

Proof. We will skip some details for brevity (see Proposition 3 of [4]). Recall that

$$\| P^t(x, \cdot) - \pi \|_{\text{TV}} = \frac{1}{2} \sum_{y \in \Omega} |P^t(x,y) - \pi(y)| \tag{4.25}$$

where $P^t(x, \cdot)$ is the distribution (at time t) of the chain which starts at x. By detailed balance the matrix $A = D^{-1}PD$ is symmetric, where D is a diagonal matrix with $D(x,x) = \sqrt{\pi(x)}$. Therefore $A = VBV^T$ for a diagonal matrix B and orthogonal V. We may then write P as $D^{-1}VBV^TD$.

Recall that from Lemma 4.11 there exists some z such that $B(z,z)=1$. V^TD are the left eigenvectors of P and so $V_w(z)$ is then $\sqrt{\pi(z)}$.
We find that

$$P^t = (D^{-1}VBV^TD)(D^{-1}VBV^TD)\dots(D^{-1}VBV^TD) = D^{-1}VB^tV^TD \tag{4.26}$$

and so

$$P^t(x,y) = \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}}\sum_w V_w(x)V_w(y)B(w,w)^t = \pi(y) + \sum_{w\neq z} V_w(x)V_w(y)\underbrace{B(w,w)^t}_{\leq(1-\lambda)^t}. \tag{4.27}$$

Returning to Equation 4.25 we obtain

$$\frac{1}{2}\sum_{y\in\Omega}|P^t(x,y)-\pi(y)| \leq \frac{(1-\lambda)^t}{2}\sum_{y\in\Omega}\left|\sum_{w\neq z}V_w(x)V_w(y)\right| \tag{4.28}$$

and $(1-\lambda)^t \leq e^{-\lambda t}$. Using Cauchy-Schwartz we may bound $\sum_{y\in\Omega}\left|\sum_{w\neq z}V_w(x)V_w(y)\right|$ by $1/\pi(x)$. Finally, when we solve for t such that $\|P^t(x,\cdot)-\pi\|_{\text{TV}}\leq\epsilon$ we obtain

$$\frac{e^{-\lambda t}}{\pi(x)} \approx \epsilon \iff -\lambda t + \log\left(\frac{1}{\pi(x)}\right) \approx \log(\epsilon) \iff t \approx \frac{1}{\lambda}\log\left(\frac{1}{\epsilon\cdot\pi(x)}\right). \tag{4.29}$$

□

Finally, we discuss the decomposition of the state space of a Markov chain (Ω,P,π) into disjoint Ω_0 and Ω_1 such that $\Omega=\Omega_0\cup\Omega_1$.

Definition 4.15. For $i\in\{0,1\}$ let $\bar{\pi}(i)=\sum_{x\in\Omega_i}\pi(x)$. Moreover, let $\bar{P}(i,j)=\bar{\pi}(i)^{-1}\pi(x)P(x,y)$ for $i,j\in\{0,1\}$. Then the *projection Markov chain* is defined as $\left(\{0,1\},\bar{P},\bar{\pi}\right)$. In the projection chain, Ω_0 and Ω_1 are the only states.

Definition 4.16. For $i\in\{0,1\}$ consider a *restriction Markov chain* $(\Omega_i,P_i,\cdot)$ where $P_i(x,y)=P(x,y)$ for $y\neq x, y\in\Omega_i$ and $P_i(x,x)=1-\sum_{z\in\Omega_i\setminus\{x\}}P(x,z)$. In the restriction chain you remain in a state x instead of transitioning outside of Ω_i.

The following theorem is due to Jerrum, Son, Tetali and Vigoda [6].

Theorem 4.17. *Let λ_i be the Poincare constant of the restriction chain for $i\in\{0,1\}$ and let $\bar{\lambda}$ be the Poincare constant of the projection chain. For any distinct $i,j\in\{0,1\}$ and $x\in\Omega_i$, if*

$$\bar{P}(i,j)=\sum_{y\in\Omega_j}P(x,y) \tag{4.30}$$

the the Poincare constant of (Ω,P,π) is at least $\min\{\bar{\lambda},\lambda_0,\lambda_1\}$.

4.1.3 An MCMC Algorithm for Sampling A Strongly Rayleigh Measure μ

The algorithm and proof in this section are due to Anari, Oveis Gharan, and Rezaei [1].
Consider a strongly Rayleigh measure $\mu:2^{[n]}\to\mathbb{R}_+$. Moreover, for $i\in[n]$ let $S-i$ be $S\setminus\{i\}$ and $S+i$ be $S\cup\{i\}$. We may then consider a Markov chain $\mathcal{M}_\mu$ with state space $\text{supp}\{\mu\}$ defined by the following algorithm:

> In state S choose an element $i\in S$ and $j\notin S$ uniformly at random and independently and let $T=S+j-i$. If $T\in\text{supp}\{u\}$ then transition to T with probability $\frac{1}{2}\min\{1,\mu(T)/\mu(S)\}$.

Because of the 1/2, we know that $\mathcal{M}$ is lazy. Moreover, we can show that μ is the stationary distribution.

Lemma 4.18. *The stationary distribution of $\mathcal{M}_\mu$ is μ.*

Proof. From Theorem 4.5 it is sufficient to show that μ satisfies the detailed balance condition. For $S, T \in \Omega = \text{supp}\{\mu\}$ assume without loss of generality that $\mu(S) < \mu(T)$. Then

$$\mu(S)P(S,T) = \mu(S)\frac{1}{2}\frac{\mu(T)}{\mu(S)} = \mu(T)\frac{1}{2} = \mu(T)\frac{1}{2}\min\left\{1, \frac{\mu(T)}{\mu(S)}\right\} = \mu(T)P(T,S). \tag{4.31}$$

□

Let C_μ be defined as

$$C_\mu = \min_{S,T \in \text{supp}\{\mu\}} \max\left\{P_\mu(S,T)P_\mu(T,S)\right\} \tag{4.32}$$

Theorem 4.19. *The Poincare constant λ of $\mathcal{M}_\mu$ is at least C_μ.*

This may be shown by induction on $\text{supp}\{\mu\}$.
The proof uses a natural decomposition of this chain and the application of Theorem 4.14.
Let m be a fixed arbitrary element in $[n]$ such that $0 < \mathbb{P}_{S\sim\mu}(m \in S) < 1$. Then, let Ω_0 and Ω_1 be the natural decomposition

$$\Omega_0 = \{S \in \text{supp}\{\mu\} \ : \ m \in S\} \tag{4.33}$$
$$\Omega_1 = \{S \in \text{supp}\{\mu\} \ : \ m \notin S\}. \tag{4.34}$$

Note that the restricted chain on Ω_0 is $\mathcal{M}_{\mu|_m}$ and the restricted chain on Ω_1 is $\mathcal{M}_{\mu|_{\overline{m}}}$. Here $\mu|_m$ is used to denote the μ conditioned on $Y_m = 1$ where $Y_m(S) = 1$ if $m \in S$. Likewise $\mu|_{\overline{m}} = \{\mu : Y_m = 0\}$.
When $\text{supp}\{\mu\} = 1$ then mixing time is $C_\mu = 1$.
Now consider the case where $\text{supp}\{\mu\} > 1$. Since $\mu|_m$ and $\mu|_{\overline{m}}$ are still strongly Rayleigh, the restricted chains $\mathcal{M}_{\mu|_m}$ and $\mathcal{M}_{\mu|_{\overline{m}}}$ are covered by the inductive hypothesis. It remains to show that $\bar{\lambda} \geq C_\mu$, where $\bar{\lambda}$ is the Poincare constant of the projection chain.
Unfortunately this is not immediate as equation 4.17 is not satisfied. Consider however a new transition kernel $\hat{P}$ with the same stationary distribution as P. Moreover, $\hat{\lambda}$ the Poincare constant of $(\Omega, \hat{P}, \mu)$ lower bounds the Poincare constant of $\mathcal{M}_\mu$ and so $\lambda \geq \hat{\lambda} \geq C_\mu$.
Anari, Oveis Gharan, and Rezaei show the following Lemma.

Lemma 4.20. *There exists $\hat{P}$ such that*

1. *$\mu(x)\hat{P}(x,y) = \mu(y)\hat{P}(y,x)$ and $\hat{P}(x,y) \leq P(x,y)$ for all $x, y \in \Omega$.*
2. *For any $i \in \{0,1\}$, $x, y \in \Omega_i$, $\hat{P}(x,y) = P(x,y)$.*
3. *The projection chain of the Poincare constant, denoted by $\bar{\hat{\lambda}}$, is bounded below by C_μ.*
4. *$\hat{P}$ satisfies Theorem 4.17.*

Note that $\hat{P}(x,y) \leq P(x,y)$ for all $x, y \in \Omega$ guarantees that $\lambda \geq \hat{\lambda}$. By (2) $\lambda_i \geq \hat{\lambda}_i$. Moreover, (4) guarantees that $\hat{\lambda} \geq \min\{\bar{\hat{\lambda}}, \hat{\lambda}_1, \hat{\lambda}_2\}$. Finally, this lemma proves Theorem 4.19 as from (3)

$$\lambda \geq \hat{\lambda} \geq \min\{\bar{\hat{\lambda}}, \hat{\lambda}_1, \hat{\lambda}_2\} \geq C_\mu. \tag{4.35}$$

Lemma 4.21. *There is a function $w_{\{x,y\}} : \Omega \times \Omega \to \mathbb{R}_+$ such that $w_{\{x,y\}} > 0$ iff $P(x,y) > 0$ and for any distinct i, j and any $x \in \Omega_j$*

$$\sum_{y \in \Omega_i} w_{\{x,y\}} = \frac{\mu(x)}{\mu(\Omega_j)} \tag{4.36}$$

Using Lemma 4.21 one may show Lemma 4.20 by constructing $\hat{P}$ as follows. For any i, j and $x \in \Omega_i$, $y \in \Omega_j$ let

$$\hat{P}(x,y) = \begin{cases} \frac{C_\mu}{\mu(x)}\mu(\Omega_i)\,\mu(\Omega_j)\,w_{\{x,y\}} & i \neq j \\ P(x,y) & i = j. \end{cases} \tag{4.37}$$

Anari, Oveis Gharan, and Rezaei check that this $\hat{P}$ indeed satisfies Lemma 4.20 and construct the function $w_{\{x,y\}}$ using a max-flow argument. Central to this argument is showing that for any $A \subseteq \Omega$,

$$\frac{\mu(N(A))}{\mu(\Omega_0)} \geq \frac{\mu(A)}{\mu(\Omega_1)} \tag{4.38}$$

where

$$N(A) = \{y \in \Omega \setminus A : \exists x \in A \text{ s.t. } P(x,y) > 0\}. \tag{4.39}$$

This may be shown with negative dependence. Consider $R \sim \mu$ and let g be a random variable which is 1 if $m \in S$ and 0 otherwise. Let f be a indicator random variable which is 1 if there exists some $T \subseteq A$ such that $T \setminus \{n\} \subseteq R$. As f and g are increasing one can use negative dependence

$$\frac{\mu(N(A))}{\mu(\Omega_0)} = \mathbb{P}_\mu[f(R) = 1 | g(R) = 0] \geq \mathbb{P}_\mu[f(R) = 1 | g(R) = 1] = \frac{\mu(A)}{\mu(\Omega_1)}. \tag{4.40}$$

In essence, the max flow argument considers a bipartite graph on Ω_0 and Ω_1. They show that a graph exists where the flow on an edge out of $x \in \Omega_1$ is $\mu(x)/\mu(\Omega_1)$ and the incoming flow on an edge into $y \in \Omega_0$ is $\mu(y)/\mu(\Omega_0)$.
By completing this argument, Anari, Oveis Gharan, and Rezaei may invoke Theorem 4.14 to upper bound the stopping time by $\frac{1}{C_\mu} \cdot \log\left(\frac{1}{\epsilon \cdot \pi(x)}\right)$.

5 Extension to a k-DPP

Anari, Oveis Gharan, and Rezaei present the following theorem.

Theorem 5.1. *Given an ensemble matrix* $\mathbf{L}$ *of a* k*-DPP* μ *and* $\epsilon > 0$ *there is an algorithm that gives an* ϵ *approximate sample in time* $\text{poly}(k)\mathcal{O}(n \log(n/\epsilon))$.

As a corollary of the argument in Section 4.1.3, the same MCMC technique may be used to achieve an upper bound

$$\tau_S(\epsilon) \leq \frac{1}{C_\mu} \cdot \log\left(\frac{1}{\epsilon \cdot \mu(S)}\right) \tag{5.1}$$

where S is a start state in $\text{supp}\{\mu\}$.
All that remains is to provide a $\mathcal{O}(n)\,\text{poly}(k)$ algorithm for sampling $S \in \text{supp}\{\mu\}$ such that $\mu(S) = \det(\mathbf{L}_S) \geq n^{-k}$, as C_μ is at least $1/2nk$ by construction.
Anari, Oveis Gharan, and Rezaei show that the greedy algorithm works for finding such S. In other words, build S by successively drawing elements from $[n]$ until $|S| = k$. At each iteration, choose the element j which maximizes $\det(\mathbf{L}_{S+j})$.

References

[1] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. Monte carlo markov chain algorithms for sampling strongly rayleigh distributions and determinantal point processes. In *Conference on Learning Theory*, pages 103–115, 2016.

[2] Julius Borcea, Petter Brändén, and Thomas Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, 2009.

[3] Petter Brändén. Polynomials with the half-plane property and matroid theory. *Advances in Mathematics*, 216(1):302–320, 2007.

[4] Persi Diaconis, Daniel Stroock, et al. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[5] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[6] Mark Jerrum, Jung-Bae Son, Prasad Tetali, Eric Vigoda, et al. Elementary bounds on poincaré and log-sobolev constants for decomposable markov chains. *The Annals of Applied Probability*, 14(4):1741–1765, 2004.

[7] Kumar Joag-Dev, Frank Proschan, et al. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295, 1983.

[8] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42. Cambridge University Press, 2017.

[9] Robin Pemantle. Towards a theory of negative dependence. *Journal of Mathematical Physics*, 41(3):1371–1390, 2000.