Chapter 16

Strongly Rayleigh Measure

We see some applications of the theory of real-stable polynomials in probability theory. Some results will be used in the next chapter to extend the method of interlacing family of polynomials.

We have mentioned in Section 13.1 that the real-rootedness of the generating polynomial of a probability distribution implies some strong properties of the probability distribution. In this chapter, we study the following generalization of this concept to the multivariate setting.

Definition 16.1 (Strongly Rayleigh Measure)**.** *Given a probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$, the generating polynomial is defined as*

$$g_\mu(x_1, \ldots, x_m) := \sum_{S \subseteq [m]} \mu(S) \prod_{i \in S} x_i.$$

We say μ is strongly Rayleigh if its generating polynomial g_μ is a real-stable polynomial.

We will see some interesting examples in Section 16.1, and some useful properties in Section 16.2. Some parts of this chapter are from the course notes of Oveis Gharan [Ove15, Ove20].

16.1 Determinantal Measure

An important example of strongly Rayleigh measure is determinantal measure. This is also called the determinantal point process in the literature (see [KT12]).

Definition 16.2 (Determinantal Measure)**.** *Let X be a random variable over $\{0,1\}^m$ with probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$. We say μ is determinantal if there exists a matrix $A \in \mathbb{R}^{m \times m}$ such that*

$$\Pr(S \subseteq X) = \sum_{R : R \supseteq S} \mu(R) = \det(A_{S,S})$$

for every subset $S \subseteq [m]$, where $A_{S,S}$ is the $|S| \times |S|$-submatrix of A restricting to the rows and columns corresponding to S.

There is a compact formula to write the generating polynomial of μ in terms of A.

Proposition 16.3 (Generating Polynomial of Determinantal Measure)**.** *If $\mu : \{0,1\}^m \to \mathbb{R}$ is determinantal with an $m \times m$ matrix $0 \preccurlyeq A \preccurlyeq I$, then the generating polynomial is*

$$g_\mu(x) = \det(I - A + A \cdot \operatorname{diag}(x)).$$

Proof. Let $h(x) = \det(I - A + A \cdot \operatorname{diag}(x))$ where $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$. We will prove that $h(\chi_S) = g_\mu(\chi_S) = \sum_{R \subseteq S} \mu(R)$ for every subset $S \subseteq [m]$ where χ_S is the characteristic vector of the subset S, and this will imply that $h(x) = g_\mu(x)$. Note that

$$h(\chi_S) = \det\left(I - A + A \cdot \operatorname{diag}(\chi_S)\right) = \det\begin{pmatrix} I_{|S|} & -A_{S,\bar{S}} \\ 0 & I_{m-|S|} - A_{\bar{S},\bar{S}} \end{pmatrix} = \det(I_{m-|S|} - A_{\bar{S},\bar{S}}).$$

Let $\det(A_{\emptyset,\emptyset}) = 1$. Recall the expansion of the characteristic polynomial in Fact 2.31 that

$$\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[n]}{k}} \det(A_{S,S}).$$

Applying this formula on h,

$$h(\chi_S) = 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R:R \subseteq \bar{S}, |R|=k} \det(A_{R,R}) = 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R:R \subseteq \bar{S}, |R|=k} \Pr(X \supseteq R)$$

where the second equality is by the definition of determinantal measure where X denotes the random outcome. Using the inclusion-exclusion principle that

$$\Pr(X \cap Y \neq \emptyset) = \sum_{k=1}^{|Y|} (-1)^{k+1} \sum_{R:R \subseteq Y, |R|=k} \Pr(X \supseteq R)$$

for a fixed subset Y and plugging in $Y = \overline{S}$, the above expression can be simplified to

$$h(\chi_S) = 1 - \Pr(X \cap \overline{S} \neq \emptyset) = \Pr(X \cap \overline{S} = \emptyset) = \Pr(X \subseteq S) = \sum_{R \subseteq S} \mu(R) = g(\chi_S).$$

□

Then the proof that determinantal measure is strongly Rayleigh follows from the results of real-stable polynomials in Chapter 13.

Theorem 16.4 (Determinantal Measure is Strongly Rayleigh). *If $\mu : \{0,1\}^m \to \mathbb{R}$ is determinantal, then μ is strongly Rayleigh.*

Proof. Using Proposition 16.3, we just need to prove that $h(x) = \det(I - A + A \cdot \operatorname{diag}(x))$ is a real-stable polynomial. We prove the claim when $0 \prec A \preccurlyeq I$, and the claim for $0 \preccurlyeq A \preccurlyeq I$ will follow from a continuity argument using Hurwitz's Theorem 13.14 as in Proposition 13.13. Note that $h(x) = \det(I - A + A \cdot \operatorname{diag}(x)) = \det(A) \cdot \det(A^{-1} - I + \operatorname{diag}(x))$, where we used that $A \succ 0$ so that A^{-1} exists and also $\det(A) > 0$. Since $0 \prec A \preccurlyeq I$, it follows that $B := A^{-1} - I \succcurlyeq 0$, and so $\det(A^{-1} - I + \operatorname{diag}(x))$ can be written as $\det(B + \sum_{i=1}^m x_i \operatorname{diag}(e_i))$ where $B \succcurlyeq 0$ and $\operatorname{diag}(e_i) \succcurlyeq 0$ for $1 \leq i \leq m$. By Proposition 13.12, $\det(B + \sum_{i=1}^m x_i \operatorname{diag}(e_i))$ is a real-stable polynomial and so is $h(x)$. (Actually, the proof in Proposition 13.12 only proves the case when $B \succ 0$, but the case $B \succcurlyeq 0$ follows from the same continuity argument using Hurwitz's Theorem 13.14.) □

Volume Measure

One interesting example of determinantal measure is the volume measure.

Definition 16.5 (Volume Measure)**.** *Given vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ that satisfies $\sum_{i=1}^m v_i v_i^T = I_n$, the volume measure μ is defined as $\mu(S) = \det\left(\sum_{i \in S} v_i v_i^T\right)$ for each subset $S \subseteq [m]$ with $|S| = n$. Note that μ is well-defined by the Cauchy-Binet formula in Fact 2.30, as*

$$1 = \det(I_n) = \det\left(\sum_{i=1}^m v_i v_i^T\right) = \sum_{S: S \subseteq [m], |S|=n} \det\left(\sum_{i \in S} v_i v_i^T\right) = \sum_{S: S \subseteq [m], |S|=n} \mu(S).$$

The following will be the base case of the proof that volume measure is determinantal.

Lemma 16.6 (Marginal Probability of Volume Measure)**.** *Let X be a random output of the volume measure μ. Then* $\Pr(j \in X) = \|v_j\|^2$.

Proof. Let V be the $n \times m$ matrix with the j-th column being v_j for $1 \leq j \leq m$. Let $V_{\bar{j}}$ be the $n \times (m-1)$ matrix where the j-th column of V is removed. By the Cauchy-Binet formula in Fact 2.30,

$$\Pr(j \notin X) = \sum_{S: j \notin S, |S|=n} \det\left(\sum_{i \in S} v_i v_i^T\right) = \det(V_{\bar{j}} V_{\bar{j}}^T).$$

By the matrix determinantal formula in Fact 2.29 and the assumption that $VV^T = I_n$,

$$\det(V_{\bar{j}} V_{\bar{j}}^T) = \det(VV^T - v_j v_j^T) = \det(VV^T)\big(1 - v_j^T (VV^T) v_j\big) = 1 - \|v_j\|^2.$$

□

The Gram matrix of the vectors $v_1, \ldots, v_m$ shows that the volume measure is determinantal. The proof of the following theorem uses that the formula for the characteristic polynomial and the inclusion-exclusion principle are the same.

Theorem 16.7 (Volume Measure is Determinantal)**.** *Let $Y = V^T V$ be the $m \times m$ Gram matrix of the vectors $v_1, \ldots, v_m \in \mathbb{R}^n$. Let X be a random output of the volume measure μ. For any $S \subseteq [m]$,*

$$\Pr_{X \sim \mu}[S \subseteq X] = \det(Y_{S,S}).$$

Proof. We prove by induction on the size of S. The base case when $|S| = 1$ was done in Lemma 16.6. For the induction step, as in the proof of Lemma 16.6, note that

$$\Pr_{X \sim \mu}[X \cap S = \emptyset] = \frac{\det\left(VV^T - \sum_{i \in S} v_i v_i^T\right)}{\det(VV^T)} = \det\left(I_n - \sum_{i \in S} v_i v_i^T\right).$$

On one hand, by the inclusion-exclusion principle,

$$\Pr_{X \sim \mu}[X \cap S = \emptyset] = 1 - \Pr_{X \sim \mu}[X \cap S \neq \emptyset] = 1 + \sum_{k=1}^{|S|} (-1)^k \sum_{R: R \subseteq S, |R|=k} \Pr_{X \sim \mu}[X \supseteq R].$$

On the other hand, let V_S be the $n \times |S|$ submatrix of V with the columns in S. By $\det(I+XY) = \det(I+YX)$ in Fact 2.28 and the formula for the characteristic polynomial in Fact 2.31,

$$\det\left(I_n - \sum_{i\in S} v_i v_i^T\right) = \det(I_n - V_S V_S^T) = \det(I_{|S|} - V_S^T V_S) = 1 + \sum_{k=1}^{|S|}(-1)^k \sum_{R:R\subseteq S, |R|=k} \det(V_R^T V_R).$$

By the induction hypothesis, $\Pr(R \subseteq X) = \det(Y_{R,R}) = \det(V_R^T V_R)$ for all $R \subset S$. So, there is a one-to-one correspondence between the (inner) summands in the above two equations for $|R| \leq k-1$, and hence we must have $\Pr_{X\sim\mu}[S \subseteq X] = \det(V_S^T V_S) = \det(Y_{S,S})$ as stated. □

Combining Theorem 16.7 and Proposition 16.3 gives a formula for the generating polynomial of the volume measure.

Corollary 16.8 (Generating Polynomial of Volume Measure)**.** *Let μ be a volume measure as defined in Definition 16.5, and Y be the Gram matrix of the vectors as defined in Theorem 16.7. Then the generating polynomial is*

$$g_\mu(x) = \det\big(I - Y + Y \cdot \operatorname{diag}(x)\big).$$

Spanning Tree Measure

An interesting consequence of Theorem 16.7 is that the uniform distribution on spanning trees is determinantal.

Definition 16.9 (Spanning Tree Measure)**.** *Let $G = (V, E)$ be an undirected graph with edge weight w_e on each edge $e \in E$. Let the edge set E be the ground set. Let $\mu : \{0,1\}^{|E|} \to \mathbb{R}$ be the probability distribution with $\mu(T) \propto \prod_{e\in T} w_e$ if T is a spanning tree and zero otherwise.*

The uniform distribution of the spanning trees in a graph is a special case when $w_e = 1$ for all $e \in E$. Using the proof in the matrix tree theorem in Problem 3.24 and the reduction to the identity matrix as in Lemma 9.11, one can show that the spanning tree measure is a volume measure.

Problem 16.10 (Burton-Pemantle Theorem)**.** *Prove that the spanning tree measure for any weighted undirected graph is a volume measure (and hence determinantal and strongly Rayleigh).*

(It may be more convenient to consider the matrix $L(G) + \vec{1}\vec{1}^T$ so that it is invertible and do the matrix tree theorem with this modified Laplacian matrix.)

A nice corollary of Problem 16.10 is that we have a nice formula from Theorem 16.7 to compute the probability that a subset of edges $F \subseteq E$ is contained in a random spanning tree.

16.2 Properties of Strongly Rayleigh Measure

Some useful properties of strongly Rayleigh measures follow from closure properties of real stable polynomials in Chapter 13.

Exercise 16.11 (Strongly Rayleigh Preserving Operations)**.** *Suppose $\mu : \{0,1\}^m \to \mathbb{R}$ is a strongly Rayleigh measure. Prove the following distributions are also strongly Rayleigh.*

1. *(Conditioning:) The conditional probability distributions* $\mu|_{x_i=0}$ *and* $\mu|_{x_i=1}$ *where the* i*-th variable is fixed to* 0 *or* 1 *for some* $1 \le i \le m$.
2. *(Projection:) For a subset* $S \subseteq [m]$*, the projection of* μ *onto* S*, denoted by* $\mu|_S$*, is the distribution supported on subsets of* S*, where for any* $R \subseteq S$*,*
$$\mu|_S(R) = \sum_{T \subseteq [m] : T \cap S = R} \mu(T).$$
3. *(External Field:) Given a non-negative vector* $(\lambda_1, \ldots, \lambda_m)$*,* $\mu * \lambda$ *is the distribution where*
$$\mu * \lambda(S) = \mu(S) \cdot \prod_{i \in S} \lambda_i.$$

The following exercise shows that some concentration property holds for strongly Rayleigh distributions.

Exercise 16.12 (Rank Sequence). *Suppose* $\mu : \{0,1\}^m \to \mathbb{R}$ *is a strongly Rayleigh measure. For* $0 \le i \le m$*, let* $a_i = \Pr_{S \sim \mu}[|S| = i]$*. Use Problem 13.7 to show that the sequence* $a_0, \ldots, a_m$ *is ultra log-concave as defined in Definition 13.6.*

Truncation

We will use the following result from real stable polynomials.

Lemma 16.13 (Homogenization). *Given a polynomial* $p \in \mathbb{R}[x_1, \ldots, x_m]$*, the homogenized version of* p*, denoted by* p_H*, is defined as*
$$p_H(x_1, \ldots, x_m, x_{m+1}) = x_{m+1}^{\deg p} \cdot p\Big(\frac{x_1}{x_{m+1}}, \ldots, \frac{x_m}{x_{m+1}}\Big).$$
For any real stable polynomial $p \in \mathbb{R}[x_1, \ldots, x_m]$ *with non-negative coefficients,* p_H *is real stable.*

The following truncation operation is quite useful.

Theorem 16.14 (Truncation). *Given a distribution* μ *and an integer* $k \ge 1$*, the truncation of* μ *is defined as the distribution* μ_k *where* $\mu_k(S) \propto \mu(S)$ *if* $|S| = k$ *and zero otherwise. For any strongly Rayleigh distribution* μ *and any* $1 \le k \le n$*,* μ_k *is strongly Rayleigh.*

Proof. Let $g_\mu(x_1, \ldots, x_m)$ be the generating polynomial of μ. As μ is strongly Rayleigh, g_μ is real stable. Consider the homogenized version $(g_\mu)_H$ of g_μ. By Lemma 16.13, $(g_\mu)_H$ is also real stable. Let the degree of g_μ be d. Observe that the generating polynomial of μ_k is
$$g_{\mu_k} \propto \partial_{x_{m+1}}^{d-k} g_{\mu H}\Big|_{x_{m+1}=0},$$
because only the terms with the degree of x_{m+1} being $d - k$ remained, and those terms have total degree exactly k in other variables $x_1, \ldots, x_m$ in the homogenized polynomial $(g_\mu)_H$. By Exercise 13.17 and Proposition 13.13, differentiation and specialization preserves real stability and so g_{μ_k} is real stable. Therefore, by Definition 16.1, μ_k is strongly Rayleigh. □

This provides an alternative proof that the volume measure is strongly Rayleigh.

Problem 16.15. *Let $L \succcurlyeq 0$ be an $m \times m$ matrix. Prove that the polynomial $\sum_{S:S\subseteq[m]} \det(L_{S,S}) \cdot z^S$ is real stable. Conclude that the volume measure in Definition 16.5 is strongly Rayleigh.*

One useful implication is that the determinantal point process restricted to size k subsets, called k-DPP, is still strongly Rayleigh, even though it is no longer determinantal. So, in particular, k-DPP still enjoys the nice properties of strongly Rayleigh distributions, including the negative correlation property in the following subsection.

Negative Correlation

This is probably the most important property of strongly Rayleigh distributions, as for instance it allows us to apply Chernoff bounds on the variables to prove concentration results.

The simplest form of negative dependency is $\Pr(x_i = 1 \mid x_j = 1) \leq \Pr(x_i = 1)$. Note that the probability $\Pr(x_i = 1)$ can be read from the generating probability as

$$\Pr(x_i = 1) = \partial_{x_i} g(x_1, \ldots, x_m)\Big|_{x_1=\ldots=x_m=1} = \sum_{S:S\ni i} \mu(S),$$

the sum of the coefficients containing i. Therefore, we can rewrite the negative correlation inequality as $\Pr(x_i = 1 \cap x_j = 1) \leq \Pr(x_i = 1) \cdot \Pr(x_j = 1)$, and then express it using generating polynomial as

$$\big(\partial_{x_i}\partial_{x_j} g(\vec{1})\big) \cdot g(\vec{1}) \leq \big(\partial_{x_i} g(\vec{1})\big)\big(\partial_{x_j} g(\vec{1})\big).$$

Strongly Rayleigh measures satisfy this inequality for any $y \in \mathbb{R}^m$, not just for $y = \vec{1}$.

Theorem 16.16 (Negative Correlation)**.** *Let $g(x_1, \ldots, x_m)$ be a multi-linear real stable polynomial. Then, for all $i \neq j$, for all $y \in \mathbb{R}^m$,*

$$\big(\partial_{x_i}\partial_{x_j} g(y)\big) \cdot g(y) \leq \big(\partial_{x_i} g(y)\big)\big(\partial_{x_j} g(y)\big).$$

Proof. For any $y \in \mathbb{R}^m$, consider the bivariate restriction

$$f(s,t) = g(y_1, \ldots, y_{i-1}, y_i + s, y_{i+1}, \ldots, y_{j-1}, y_j + t, y_{j+1}, \ldots, y_m).$$

Then f is a bivariate real stable polynomial. Since g is multi-linear, note that

$$f(s,t) = g(y) + s \cdot \partial_{x_i} g(y) + t \cdot \partial_{x_j} g(y) + st \cdot \partial_{x_i}\partial_{x_j} g(y).$$

The univariate polynomial $h(s) = f(s, \imath)$ is $\mathcal{H}$-stable (but not necessarily real). Let $a + \imath b$ be a root of $h(s)$. Then

$$\Re(h(a + \imath b)) = g(y) + a \cdot \partial_{x_i} g(y) - b \cdot \partial_{x_i}\partial_{x_j} g(y) = 0,$$

$$\Im(h(a + \imath b)) = b \cdot \partial_{x_i} g(y) + \partial_{x_j} g(y) + a \cdot \partial_{x_i}\partial_{x_j} g(y) = 0.$$

Solving the two equations by eliminating a, we get

$$\big(\partial_{x_i}\partial_{x_j} g(y)\big) \cdot g(y) - \big(\partial_{x_i}\partial_{x_j} g(y)\big)^2 \cdot b = \big(\partial_{x_i} g(y)\big)^2 \cdot b + \big(\partial_{x_i} g(y)\big) \cdot \big(\partial_{x_j} g(y)\big).$$

By stability of h, we have $b \leq 0$. This implies that $(\partial_{x_i}\partial_{x_j} g(y)) \cdot g(y) - (\partial_{x_i} g(y)) \cdot (\partial_{x_j} g(y)) \leq 0$. ☐

The converse of Theorem 16.16 is also true; see [Ove20].

Negative Association

A stronger form of negative dependency is called negative association.

Definition 16.17 (Negative Association)**.** *The binary random variables* $\{x_1, \ldots, x_m\}$ *are negatively associated if for any two non-decreasing functions* $f, g \in \{0,1\}^m \to \mathbb{R}$ *that depend on disjoint set of variables, it holds that*

$$\mathbb{E}\left[f(x_1, \ldots, x_m) \cdot g(x_1, \ldots, x_m)\right] \leq \mathbb{E}\left[f(x_1, \ldots, x_m)\right] \cdot \mathbb{E}\left[g(x_1, \ldots, x_m)\right],$$

where a function f *is nondecreasing if* $f(\vec{x}) \geq f(\vec{y})$ *if* $\vec{x} \geq \vec{y}$*.*

Note that negative correlation is a special case of negative association. Feder and Mihail [FM92] used negative correlation as the base case in an induction to prove that the random variables of a strongly Rayleigh measure are negatively associated. Borcea, Brändén and Liggett [BBL09] developed the theory of strongly Rayleigh measure and use it to answer many questions about negatively dependent random variables (see [Pem11]).

One consequence of negative association is that Chernoff bounds apply on the random variables from a strongly Rayleigh distribution, even though they are not independent.

Theorem 16.18 (Concentration of Strongly Rayleigh Distribution)**.** *Let* $\mu : \{0,1\}^m \to \mathbb{R}$ *be a* k*-homogeneous strongly Rayleigh distribution. Let* $f : \{0,1\}^m \to \mathbb{R}$ *be a* 1*-Lipchitz function where* $|f(S_1) - f(S_2)| \leq 1$ *for any two sets* $S_1, S_2 \subseteq [m]$ *with* $|S_1 \triangle S_2| = 1$*. Then, for any* $a \geq 0$*,*

$$\Pr\left[|f - \mathbb{E}[f]| > a\right] \leq e^{-\frac{a^2}{k}}.$$

Oveis Gharan used this theory to prove many interesting properties of random spanning trees, and used these properties to design and analyze approximation algorithms for traveling salesman problems. We refer the reader to his notes [Ove15] for some interesting examples.

16.3 References

[BBL09] 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. 159

[FM92] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 26–38. ACM, 1992. 159

[KT12] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Found. Trends Mach. Learn.*, 5(2-3):123–286, 2012. 153

[Ove15] Shayan Oveis Gharan. Course notes of "Recent developments in approximation algorithms", lecture 10–12. 2015. `https://homes.cs.washington.edu/~shayan/courses/approx/index.html`. 153, 159

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 4. 2020. `https://homes.cs.washington.edu/~shayan/courses/polynomials/`. 128, 130, 153, 158

[Pem11] Robin Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. *Current Developments in Mathematics*, 2011(1), 2011. 159