

Lecture 1 and 2: Random Spanning Trees

Lecturer: Shayan Oveis Gharan

March 31st

Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications.

In these lectures we prove the matrix tree theorem and Burton, Pemantle theorem.

1.1 Overview of Eigenvalues of Symmetric Matrices

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Recall that λ is an eigenvalue of A with corresponding eigenvector x , if

$$Ax = \lambda x.$$

It is easy to see that if x_1 is an eigenvector of λ_1 and x_2 is an eigenvector of λ_2 and $\lambda_1 \neq \lambda_2$, then x_1 is orthogonal to x_2 ,

$$\langle x_1, x_2 \rangle = 0.$$

Even if there is a multiplicity of eigenvalues, i.e., $\lambda_1 = \lambda_2$, we still assume that we have an eigenvector corresponding to each eigenvalue such that for any $i \geq j$, x_i is orthogonal to x_j . To justify that, notice that when we have eigenvalues with multiplicity k it means that there is a corresponding linear space of dimension k corresponding to those eigenvalues. So we can choose k orthonormal vectors in that linear space and assign each of them to one of those k equal eigenvalues.

The variational characterization of eigenvalues gives a way of estimating eigenvalues as solutions of an optimization problem. If $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are the eigenvalues of A , then

$$\begin{aligned} \lambda_k &= \min_{x_1, \dots, x_k \text{ orthogonal}} \max_{x \neq 0} \left\{ \frac{x^\top A x}{x^\top x} : x \in \text{span}\{x_1, \dots, x_k\} \right\} \\ &= \max_{x_1, \dots, x_{n-k+1} \text{ orthogonal}} \min_{x \neq 0} \left\{ \frac{x^\top A x}{x^\top x} : x \in \text{span}\{x_1, \dots, x_{n-k+1}\} \right\}. \end{aligned} \quad (1.1)$$

$x^\top A x$ is usually considered as the *natural quadratic form* of the matrix A . Perhaps one of the main uses of this characterization is to prove bounds on the eigenvalues without actually computing them.

Recall that all eigenvalues of every symmetric matrix are real. By spectral theorem, for any symmetric matrix A with real eigenvalues $\lambda_1, \dots, \lambda_n$ and corresponding eigenvectors x_1, \dots, x_n we can write,

$$A = \sum_{i=1}^n \lambda_i x_i x_i^\top.$$

Using the above characterization we can define $f(A)$ for an arbitrary function f , just by acting f on the eigenvalues of A .

$$f(A) := \sum_{i=1}^n f(\lambda_i) x_i x_i^\top.$$

For example, let $f(t) = 1/t$ if $t \neq 0$ and $f(t) = 0$ otherwise. Then we can define the *pseudo-inverse* of A as follows:

$$A^\dagger := \sum_{i: \lambda_i \neq 0} \frac{1}{\lambda_i} x_i x_i^\top.$$

As a simple application of the spectral theorem we prove (1.1) for $k = 1$. For a symmetric matrix A with eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$ we show that

$$\lambda_1 = \min_{y \neq 0} \frac{y^T A y}{y^T y}$$

Let x_1, \dots, x_n be orthonormal eigenvalues corresponding to $\lambda_1, \dots, \lambda_n$. Since x_1, \dots, x_n form an orthonormal basis we can write

$$y = \sum_{i=1}^n \langle x_i, y \rangle x_i.$$

Therefore,

$$\begin{aligned} \frac{y^T A y}{y^T y} &= \frac{y^T \left(\sum_{i=1}^n \lambda_i x_i x_i^T \right) y}{\left(\sum_{i=1}^n \langle x_i, y \rangle x_i^T \right) \cdot \left(\sum_{i=1}^n \langle x_i, y \rangle x_i \right)} \\ &= \frac{\sum_{i=1}^n \lambda_i \langle x_i, y \rangle^2}{\sum_{1 \leq i, j \leq n} \langle x_i, y \rangle \cdot \langle x_j, y \rangle \cdot \langle x_i, x_j \rangle} \\ &= \frac{\sum_{i=1}^n \lambda_i \langle x_i, y \rangle^2}{\sum_{i=1}^n \langle x_i, y \rangle^2} \end{aligned}$$

In the last equality we are using that x_1, \dots, x_n , so $\langle x_i, x_j \rangle = 1$ if $i = j$ and it is zero otherwise. Suppose $\lambda_2 > \lambda_1$. Then, obviously the above ratio is equal to λ_1 if and only y is parallel to x_1 and orthogonal to all other eigenvectors. In general, the above ratio is equal to λ_1 if and only if y is in the linear space of the eigenvectors of λ_1 .

1.2 Overview of PSD matrices

We say a symmetric matrix $A \in \mathbb{R}^{V \times V}$ is PSD if for any vector $x \in \mathbb{R}^V$,

$$x^T A x \geq 0.$$

We use the notation $A \succeq 0$ to denote A is PSD. Also, we write $A \succeq B$ to denote $A - B \succeq 0$.

Next, we say equivalent characterization of PSD matrices.

- $A \succeq 0$ iff all eigenvalues of A are non-negative. This can be proved easily using the variational characterization (1.1).
- We say a matrix B is rank one, if $B = x^T y$ for $x, y \in \mathbb{R}^V$. $A \succeq 0$ iff there are vectors x_1, \dots, x_d , where d is the rank of A such that

$$A = \sum_{i=1}^d x_i x_i^T.$$

- $A \succeq 0$ iff for any set $S \subseteq [n]$, $A_S \succeq 0$. This is because we can extend any vector $x_S \in \mathbb{R}^S$, to $x \in \mathbb{R}^V$ where

$$x(v) = \begin{cases} x_S(v) & \text{if } v \in S \\ 0 & \text{otherwise} \end{cases}.$$

Then,

$$x_S^T A_S x_S = x^T A x \geq 0.$$

If $A \succeq 0$ then we can define the square root of A , $A^{1/2}$ as follows:

$$A^{1/2} := \sum_{i: \lambda_i > 0} \sqrt{\lambda_i} x_i x_i^T.$$

1.3 The Laplacian Matrix

Throughout this lecture we assume $G = (V, E)$ is a simple graph. Unless otherwise specified, we assume G is unweighted.

Given a graph $G = (V, E)$, fix an arbitrary orientation. For an edge $e = (u, v)$, i.e., oriented from u to v , let

$$b_e = \mathbf{1}_u - \mathbf{1}_v,$$

where

$$\mathbf{1}_u(v) = \begin{cases} 1 & \text{if } v = u, \\ 0 & \text{otherwise.} \end{cases}$$

Say, $E = \{e_1, \dots, e_m\}$, we write

$$B = \begin{pmatrix} b_{e_1} \\ b_{e_2} \\ \vdots \\ b_{e_m} \end{pmatrix}$$

to denote the matrix in $\mathbb{R}^{E \times V}$ with rows correspond to the edges of G .

The Laplacian of an edge $e = \{u, v\}$ is defined as

$$L_{\{u,v\}} = b_{\{u,v\}} b_{\{u,v\}}^\top.$$

It is a matrix that is equal to 1 in the u, u and v, v entries and -1 in the u, v and v, u entries and 0 otherwise.

The Laplacian of G , L_G is defined as follows:

$$L_G := \sum_{e \in E} b_e b_e^\top = B^\top B.$$

If the graph is clear in the context we may drop the subscript G . Equivalently, we can define the Laplacian of G as

$$L_G = D - A,$$

where D is the diagonal matrix of vertex degrees and A is the adjacency matrix. If G is a weighted graph, say $w(e)$ is the weight of each edge e , then we write

$$L_G = \sum_{e \in E} w(e) b_e b_e^\top.$$

It is also instructive to define the Laplacian of the subgraphs of G . For a set $T \subseteq E$, we let

$$L_T := \sum_{e \in T} b_e b_e^\top.$$

Perhaps the most natural property of the Laplacian matrix is the simple description of their natural quadratic form. For any function $f : V \rightarrow \mathbb{R}$,

$$x^\top L_G x = \sum_{u,v \in E} |x(u) - x(v)|^2.$$

It follows that for any graph G , the Laplacian of G is PSD, $L_G \succeq 0$. In addition, at least one of the eigenvalues of L_G is zero.

One simple proof of the above identity is to observe that for any edge $e = \{u, v\}$,

$$x^\top L_e x = |x(u) - x(v)|^2.$$

Then,

$$x^\top L_G x = x^\top \left(\sum_{e \in E} L_e \right) x = \sum_{e \in E} x^\top L_e x.$$

In almost all parts of this course we work with directed graph, so we can assume that L_G has exactly one zero eigenvalue. It is sometimes instructive to work with

$$\tilde{L}_G := L_G + \frac{1}{n} J,$$

where J is simply the all-1s matrix. Since \tilde{L}_G is non-singular we can analytically study and use its determinant.

1.4 Properties of the Determinant

For a square matrix A we write $\det(A)$ to determinant of A .

Fact 1.1. For any two square matrices A, B ,

$$\det(AB) = \det(A) \det(B).$$

Also, for any matrix A ,

$$\det(A) = \det(A^\top).$$

We leave these as exercises.

Fact 1.2. For any symmetric matrix A with eigenvalues $\lambda_1, \dots, \lambda_n$,

$$\det(A) = \prod_{i=1}^n \lambda_i.$$

The next is the very useful tool known as the Cauchy-Binet formula.

Lemma 1.3 (Cauchy-Binet). Let $x_1, \dots, x_m \in \mathbb{R}^n$ and $y_1, \dots, y_m \in \mathbb{R}^n$. Then,

$$\det \left(\sum_{i=1}^m x_i y_i^\top \right) = \sum_{S \in \binom{[m]}{n}} \det \left(\sum_{i \in S} x_i y_i^\top \right).$$

Equivalently, say $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$ for $m \geq n$, we can write

$$\det_n(AB) = \det_n(BA).$$

Given a set of vectors x_1, \dots, x_k , the k -dimensional parallelepiped is the set of points that are convex combinations of these points and the zero, i.e.,

$$\{\alpha_1 x_1 + \dots + \alpha_k x_k : 0 \leq \alpha_i \leq 1, 1 \leq i \leq k\}.$$

The following theorem is immediate.

Theorem 1.4. Given a set of vectors $x_1, \dots, x_k \in \mathbb{R}^d$, let

$$A = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$$

The k -dimensional volume of the k -parallelepiped P defined on x_1, \dots, x_k is equal $\sqrt{\det(AA^\top)}$,

$$\text{vol}(P)^2 = \det(AA^\top).$$

As an application application of the theorem we can upper-bound $\det(AA^\top)$.

Corollary 1.5. For any matrix $A \in \mathbb{R}^{k \times d}$ with rows x_1, \dots, x_k ,

$$\det(AA^\top) \leq \prod_{i=1}^k \|x_i\|^2.$$

1.5 Counting Spanning Trees

Perhaps one of the earliest applications of the Laplacian matrix was to count the number of spanning trees of G . Kirchoff proved the following theorem.

Theorem 1.6 (Matrix Tree Theorem). For any graph G , the number of spanning trees of G is equal to

$$\frac{1}{n} \det(\tilde{L}_G).$$

More generally, we can study the sum of all weighted spanning trees of G . Let $w : E \rightarrow \mathbb{R}_+$ be a non-negative weights assigned to the edges of G . Then, the determinant of the weighted Laplacian of G is equal to the sum all weighted spanning trees of G , i.e.,

$$\det\left(\mathbf{1}\mathbf{1}^\top + \sum_{e \in E} w(e)b_e b_e^\top\right) = n \sum_{T \in \mathcal{T}} \prod_{e \in T} w(e).$$

The proof of the above equality is almost similar to the proof of the Matrix Tree Theorem. Next, we prove the Matrix Tree Theorem.

Let $\mathbf{1} \in \mathbb{R}^V$ where $\mathbf{1}(v) = 1/\sqrt{n}$ for all $v \in V$. Then, by Cauchy-Binet,

$$\det\left(\mathbf{1}\mathbf{1}^\top + \sum_{e \in E} b_e^\top b_e\right) = \sum_{F \subseteq \binom{E}{n-1}} \det\left(\mathbf{1}\mathbf{1}^\top + \sum_{e \in F} b_e^\top b_e\right).$$

Now, for a set $F \subseteq E$ of size $n-1$ let $\tilde{B}_F \in \mathbb{R}^{n \times n}$ be the matrix where the first row $\mathbf{1}$ and the rest of the rows are b_e for all $e \in F$. Then,

$$\det\left(\mathbf{1}\mathbf{1}^\top + \sum_{e \in F} b_e^\top b_e\right) = \det(\tilde{B}_F \tilde{B}_F^\top).$$

We show that $\det(\tilde{B}_F \tilde{B}_F^\top) = n$ if the edges in F induce a spanning tree and it is zero otherwise. This completes the proof of the matrix tree theorem.

First, since $\mathbf{1}$ is orthogonal to b_e for any $e \in E$,

$$\tilde{B}_F \tilde{B}_F^\top = \begin{bmatrix} 1 & 0 \dots 0 \\ 0 & \\ \vdots & B_F B_F^\top \\ 0 & \end{bmatrix}$$

So, $\det(\tilde{B}_F \tilde{B}_F^\top) = \det(B_F B_F^\top)$.

Since $|F| = n - 1$, if F is not a tree it is disconnected, so \tilde{L}_F has a zero eigenvalue. This means that $\det(\tilde{B}_F \tilde{B}_F^\top) = \det(\tilde{L}_F) = 0$.

Now, suppose F is a spanning tree. We show $\det(B_F B_F^\top) = n$.

Claim 1.7. *If $F \subseteq E$ is a spanning tree, then $\det(B_F B_F^\top) = n$.*

Proof. By Cauchy-Binet,

$$\det(B_F B_F^\top) = \sum_{S \in \binom{V}{n-1}} \det(B_{F,S} B_{F,S}^\top).$$

where we use $B_{F,S}$ to denote the $\mathbb{R}^{n-1 \times n-1}$ submatrix of B_F indexed by S . Finally, we show that $|\det(B_{F,S})| = 1$ for all n choices of S .

Since F is a tree we can reorder the columns of $B_{F,S}$ such that the resulting matrix is lower-triangular. We start by finding a leaf and we let the corresponding column be the last column of $B_{F,S}$ and we recurse. The resulting matrix is lower-triangular and the diagonal entries are either $+1$ or -1 . It is very easy to see that the determinant of any lower-triangular (or upper-triangular) matrix is the product of the entries in the diagonal. So, $|\det(B_{F,S})| = 1$. \square

This completes the proof of [Theorem 1.6](#).

1.6 Random Spanning Tree Distributions

A *uniform* spanning tree distribution is a uniform distribution over all spanning trees of a given graph. In the rest of the lecture we will prove many properties of these distributions.

If the graph is weighted, then we can study the weighted uniform distribution of spanning trees where the probability of each tree is proportional to the product of the weight of its edges.

Definition 1.8 (*w*-uniform spanning tree distributions). *For $w : E \rightarrow \mathbb{R}_+$, we say μ is a *w*-uniform spanning tree distribution, if for any spanning tree $T \in \mathcal{T}$,*

$$\mathbb{P}[T] \propto \prod_{e \in T} w(e).$$

See [Figure 1.1](#) for an example of a λ -uniform spanning tree distribution. As we will see almost all of the properties of uniform spanning tree distributions naturally extend to the weighted uniform distributions.

Since we can count spanning trees, we can sample from this distribution efficiently. First, of all we can compute the probability that each edge e is in a random spanning tree. Say μ is a uniform distribution on

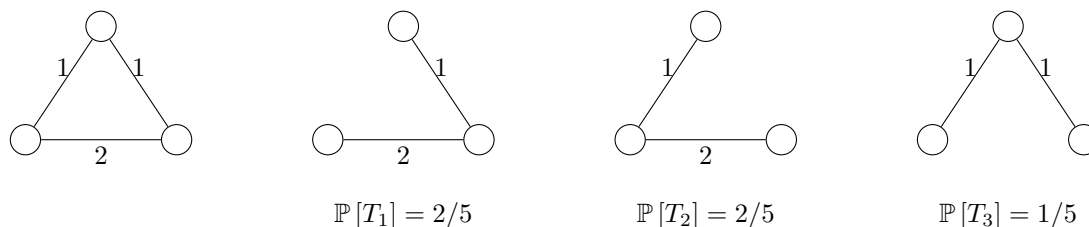


Figure 1.1: A weighted uniform spanning tree distribution

all spanning trees of G . Then,

$$\mathbb{P}_{T \sim \mu} [e \in T] = 1 - \mathbb{P}_{T \sim \mu} [e \notin T] = 1 - \frac{|\{T \in \mathcal{T} : e \notin T\}|}{|\mathcal{T}|} = 1 - \frac{\det(\tilde{L}_{G \setminus \{e\}})}{\det(\tilde{L}_G)}.$$

Once we can calculate the probability of an edge being in a random spanning tree we can sample a tree from μ efficiently. The following algorithm runs in roughly $O(n^5)$ and samples a uniform spanning tree in a given graph [Gué83].

Algorithm 1 Sampling a Uniform Spanning Tree

Let $G_0 = G$, and $T = \{\}$.

Choose an arbitrary ordering of the edges, e_1, \dots, e_m .

for $i = 1 \rightarrow m$ **do**

 Let p be the probability that e_i is in a random spanning tree of G_{i-1} .

 With probability p , add e_i to T and let $G_i = G_{i-1} / \{e_i\}$ (i.e., contract e_i).

 With the remaining probability, let $G_i = G_{i-1} \setminus \{e_i\}$ (i.e., delete e_i).

end for

1.7 Marginal Probability of Edges

Let μ be a uniform distribution of all spanning trees of G . Let $\mu_e := \mathbb{P}_{T \sim \mu} [e \in T]$, be the marginal probability of edge e . In this section we write analytical expression for μ_e .

Lemma 1.9. For any edge e ,

$$\mu_e = b_e^\top L_G^\dagger b_e.$$

More generally, if $w(e)$ is the weight of e , then

$$\mu_e = w(e) b_e^\top L_G^\dagger b_e.$$

Next, we use the determinant rank one update formula to prove the above lemma.

Theorem 1.10. For a PD matrix $A \in \mathbb{R}^{n \times n}$ and any vectors $x, y \in \mathbb{R}^n$,

$$\det(A + xx^\top) = \det(A)(1 + x^\top A^{-1}x).$$

Proof. First, by [Fact 1.1](#),

$$\det(A + x^\top x) = \det(A^{1/2}(I + A^{-1/2}x^\top A^{-1/2})A^{1/2}) = \det(A) \det(I + A^{-1/2}x^\top A^{-1/2}).$$

It is easy to see $I + A^{-1/2}x^\top A^{-1/2}$ has $n - 1$ eigenvalues that are 1 and 1 eigenvalue that is $1 + x^\top A^{-1}x$. In words for any vector y , a rank 1 update, $I + yy^\top$, only shifts one of the eigenvalues of I by $\langle y, y \rangle$. Therefore,

$$\det(I + A^{-1/2}x^\top A^{-1/2}) = 1 + x^\top A^{-1}x,$$

that completes the proof. \square

Proof of [Lemma 1.9](#). We prove the statement in the case of unweighted graphs and we leave the weighted version as an exercise.

$$\mu_e = 1 - \mathbb{P}[e \notin T] = 1 - \frac{\det(\tilde{L}_G - b_e b_e^\top)}{\det(\tilde{L}_G)}$$

By [Theorem 1.10](#), we can write

$$\mu_e = 1 - \frac{\det(\tilde{L}_G)(1 - b_e^\top \tilde{L}_G^{-1} b_e)}{\det(\tilde{L}_G)} = b_e^\top L_G^\dagger b_e,$$

where we used that $\tilde{L}_G b_e = L_G b_e$. \square

The quantity $\text{Reff}(e) := b_e^\top L_G^\dagger b_e$ is also known as the *effective resistance* of edge e . We will talk about effective resistance in more details in the next lecture. Right now, it is an immediate consequence of [Fact 1.1](#) that the sum of the effective resistances of all edges of G is exactly equal to $n - 1$.

Lemma 1.11. *For any unweighted graph $G = (V, E)$ the sum of the effective resistance of all edges is equal to $n - 1$.*

Proof. We give two proofs. Let μ be the uniform spanning tree distribution. By linearity of expectation,

$$\sum_{e \in E} \text{Reff}(e) = \sum_{e \in E} \mu_e = \mathbb{E}_{T \sim \mu} [|T|] = n - 1.$$

The above proof uses several machineries, Matrix tree theorem, rank 1 update formula, etc. The second proof uses just the spectral definition of effective resistance.

$$\begin{aligned} \sum_{e \in E} \text{Reff}(e) &= \sum_{e \in E} \text{Tr}(b_e^\top L_G^\dagger b_e) \\ &= \sum_{e \in E} \text{Tr}\left(b_e b_e^\top L_G^\dagger\right) \end{aligned}$$

In the second equality we used the identity that for matrices $A \in \mathbb{R}^{k \times n}$ and $B \in \mathbb{R}^{n \times k}$, $\text{Tr}(AB) = \text{Tr}(BA)$. Trace and sum commute, so

$$\begin{aligned} \sum_{e \in E} \text{Reff}(e) &= \text{Tr}\left(\sum_{e \in E} b_e b_e^\top L_G^\dagger\right) \\ &= \text{Tr}\left(I_{n-1} L_G^\dagger\right) \\ &= \text{Tr}(I_{n-1}) = n - 1. \end{aligned}$$

In the above equations we assumed I_{n-1} is the identity matrix in the linear space of vectors orthogonal to the all 1 vector. In particular, you can assume $I_{n-1} = I - \mathbf{1}\mathbf{1}^\top$. In the last equality we used the $\text{Tr}(A)$ is the sum of the eigenvalues of A . The last equality follows by the fact that I_{n-1} has $n - 1$ eigenvalues that are 1 and one eigenvalue that is 0. \square

1.8 Isotropic Vectors

Definition 1.12. A set of vectors $y_1, \dots, y_m \in \mathbb{R}^n$ are in isotropic position if for any unit vector $x \in \mathbb{R}^n$,

$$\sum_{i=1}^m \langle x, y_i \rangle^2 = 1.$$

Observe that we can rewrite the above equality as follows. For any unit vector x ,

$$\begin{aligned} \sum_{i=1}^m \langle x, y_i \rangle^2 &= \sum_{i=1}^m x^\top y_i y_i^\top x \\ &= x^\top \left(\sum_{i=1}^m y_i y_i^\top \right) x = 1. \end{aligned}$$

In other words, the above identity holds if and only if $\sum_{i=1}^m y_i y_i^\top = I$. Put it in another way, a set of vectors are in isotropic position if every direction in the space looks like every other direction, that is for every direction x if we project all of the vectors onto x the sum of the squares of the projections is invariant over the choices of x .

There is a natural linear transformation that can make any given set of vectors $\{b_1, \dots, b_m\}$ isotropic.

$$y_i = \left(\sum_{i=1}^m b_i b_i^\top \right)^{\dagger/2} b_i.$$

Then,

$$\begin{aligned} \sum_{i=1}^m y_i y_i^\top &= \sum_{i=1}^m \left(\sum_{i=1}^m b_i b_i^\top \right)^{\dagger/2} b_i b_i^\top \left(\sum_{i=1}^m b_i b_i^\top \right)^{\dagger/2} \\ &= \left(\sum_{i=1}^m b_i b_i^\top \right)^{\dagger/2} \left(\sum_{i=1}^m b_i b_i^\top \right) \left(\sum_{i=1}^m b_i b_i^\top \right)^{\dagger/2} = I. \end{aligned}$$

Geometrically, this transformation equalizes the eigenvalues of the matrix $\sum_{i=1}^m b_i b_i^\top$. In the next section, we use this transformation to make $\{b_e\}_{e \in E}$ vectors isotropic.

1.9 Isotropic Transformation of b_e vectors

Consider the $n - 1$ dimensional space with all of the vectors $\{b_e\}_{e \in E}$ (that is the space of vectors orthogonal to $\mathbf{1}$). In [Claim 1.7](#) we show that for any set $F \subseteq E$ that is a spanning tree, $\det(B_F B_F^\top) = n - 1$. By [Theorem 1.4](#) $\det(B_F B_F^\top)$ is the square of the volume of the parallelepiped

$$\left\{ \sum_{e \in F} \alpha_e b_e : 0 \leq \alpha_e \leq 1, \forall e \in F \right\}.$$

Say a set of $n - 1$ vectors is a basis if they are linearly independent. It follows that the set vectors $\{b_e\}_{e \in F}$ form a basis if and only if F is a spanning tree. So, basically the above argument says the volume of every

basis in this $n - 1$ dimensional space is equal to \sqrt{n} . On the other hand, by Cauchy-Binet $\det(\tilde{L}_G)$ is just the sum of the volume of all bases. That is the reason that $\det(\tilde{L}_G)$ counts the number of spanning trees of G (up to normalization).

Let us define

$$y_e := L_G^{\dagger/2} b_e. \quad (1.2)$$

If G is weighted, then we define

$$y_e := \sqrt{w(e)} \cdot L_G^{\dagger/2} b_e.$$

It follows from the discussion in the previous section that $\{y_e\}_{e \in E}$ are isotropic.

Now, let us calculate the volume of a basis after the isotropic transformation of a tree $F = \{b_{e_1}, \dots, b_{e_{n-1}}\}$. Let

$$\mathcal{B}_F = L_G^{\dagger/2} B_F = \begin{pmatrix} L_G^{\dagger/2} b_{e_1} \\ L_G^{\dagger/2} b_{e_2} \\ \vdots \\ L_G^{\dagger/2} b_{e_{n-1}} \end{pmatrix}.$$

Then,

$$\begin{aligned} \det(\mathcal{B}_F \mathcal{B}_F^T) &= \det(\tilde{L}_G^{\dagger/2} B_F B_F^T \tilde{L}_G^{\dagger/2}) \\ &= \det(\tilde{L}_G^{\dagger}) \det(B_F B_F^T) = \frac{1}{|\mathcal{T}|}. \end{aligned}$$

The last equality uses the matrix tree theorem. So, the square of the volume of each basis in the isotropic position is exactly the probability that the basis is chosen in a uniform distribution of bases.

On the other hand, for any edge e , the probability that e is in a random spanning tree is the same as the 1 dimensional volume of y_e .

$$\det(y_e y_e^T) = b_e L_G^{\dagger} b_e = \mathbb{P}_{T \sim \mu} [e \in T].$$

So, one might expect that this is always true for any arbitrary set F in the isotropic mapping. This is indeed true and it is proved by Burton and Pemantle as we will see in the next section.

1.10 Joint Probability of Subsets of Edges

In this section we show that (weighted) uniform spanning tree distributions are determinantal.

Definition 1.13 (Determinantal Distributions). *A probability distribution $\mu : 2^E \rightarrow \mathbb{R}_+$ defined on the subsets of a set E of elements is called a determinantal probability measure if there exists a matrix $Y \in \mathbb{R}^{E \times E}$ such that for any set $F \subseteq E$,*

$$\mathbb{P}_{T \sim \mu} [F \in T] = \det(Y_F),$$

where Y_F is the principal submatrix of Y indexed by the rows and columns corresponding to F .

Perhaps the simplest examples of the determinantal distributions are distributions of independent Bernoulli random variables. Say B_1, \dots, B_n are independent Bernoulli random variables where B_i occurs with probability p_i . Then, we can simply let $Y \in \mathbb{R}^{n \times n}$ be a diagonal matrix where $Y(i, i) = p_i$.

Next, we discuss a beautiful theorem of Burton and Pemantle [BP93] that shows that random spanning tree distributions are families of determinantal measures.

Theorem 1.14 (Burton and Pemantle [BP93]). *Given a graph G , let $Y \in \mathbb{R}^{E \times E}$ where for each pair of edges $e, f \in E$,*

$$Y(e, f) = \langle y_e, y_f \rangle.$$

Then, for any set of edges $F \subseteq E$,

$$\mathbb{P}_{T \sim \mu} [F \subseteq T] = \det(\mathcal{B}_F \mathcal{B}_F^\top) = \det(Y_F).$$

It also follows from the above theorem of Burton and Pemantle that λ -uniform spanning tree distributions are determinantal measures. We will see a generalization of determinantal measures in the future lectures. We will prove the above theorem in the next lecture. Right now, we discuss some of its implications.

An immediate consequence of [Theorem 1.14](#) is that each pair of edges are negatively correlated.

Fact 1.15 (Pairwise Negative Correlation). *For any pair of edges $e, f \in E$,*

$$\mathbb{P}_{T \sim \mu} [e, f \in T] \leq \mathbb{P}[e \in T] \cdot \mathbb{P}[f \in T].$$

To see the proof, by Burton, Pemantle,

$$\begin{aligned} \mathbb{P}[e, f \in T] &= \det \begin{pmatrix} \langle y_e, y_e \rangle & \langle y_e, y_f \rangle \\ \langle y_e, y_f \rangle & \langle y_f, y_f \rangle \end{pmatrix} \\ &= \|y_e\|^2 \cdot \|y_f\|^2 - \langle y_e, y_f \rangle \langle y_f, y_e \rangle. \end{aligned}$$

Using [Lemma 1.9](#),

$$\mathbb{P}[e, f \in T] - \mathbb{P}[e \in T] \cdot \mathbb{P}[f \in T] = -\langle y_e, y_f \rangle^2 \leq 0.$$

So, e, f are negatively correlated. But, we can even analytically write down the correlation between each pair of edges. For example, if $\langle y_e, y_f \rangle = 0$ then we can say edge e is independent of edge f .

More generally, we can define negative correlation of a subset of edges.

Definition 1.16 (Negative Correlation). *For a distribution $\mu : 2^E \rightarrow \mathbb{R}_+$, we say a set of edges $F \subseteq E$, are negatively correlated if*

$$\mathbb{P}_{T \sim \mu} [F \subseteq T] \leq \prod_{e \in F} \mathbb{P}[e \in T].$$

It follows by [Theorem 1.14](#) and [Corollary 1.5](#) that in any random spanning tree distribution any set of edges are negatively correlated. In particular, the LHS of the above equation is the square of the volume of the $|F|$ -parallelepiped defined on $\{y_e\}_{e \in F}$ and the RHS is just $\prod_{e \in F} \|y_e\|^2$.

References

- [BP93] R. Burton and R. Pemantle. “Local Characteristics, Entropy and Limit Theorems for Spanning Trees and Domino Tilings via Transfer-Impedances”. In: *Ann. Probab.* 21.math.PR/0404048 (1993), pp. 1329–1371 (cit. on pp. [1-10](#), [1-11](#)).
- [Gué83] A. Guénoche. “Random Spanning Tree”. In: *J. Algorithms* 4 (1983), pp. 214–220 (cit. on p. [1-7](#)).