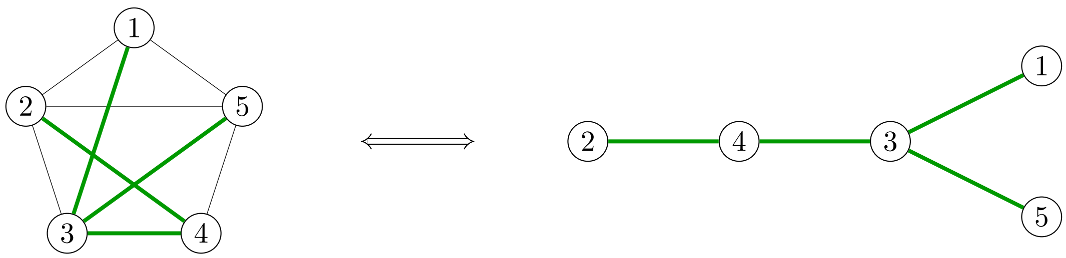
Math 374
Matrix Tree Theorem

1 Counting spanning trees: A determinantal formula

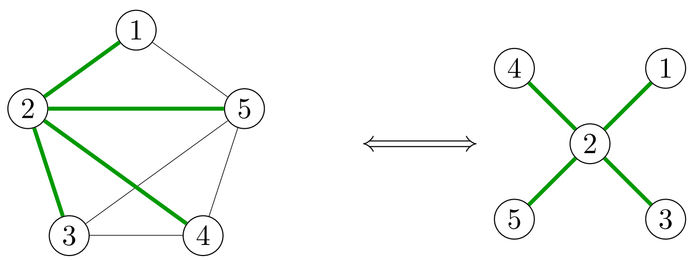
Recall that a *spanning tree* of a graph G is a subgraph T so that T is a tree and $V(G) = V(T)$.

Question. How many distinct spanning trees are there in an arbitrary graph?

If we set $\tau(G)$ to be the number of spanning trees in a graph G, then we actually already have an expression, thanks to Cayley's formula, for $\tau(K_n)$. To see this label the vertices of K_n with the labels $1, 2, \ldots, n$. Now observe that any spanning tree in K_n is just a (vertex) labeled tree with order n and vice versa. Therefore, $\tau(K_n) = n^{n-2}$. To see this in action consider two of the $5^3 = 125$ spanning trees in K_5:



and



Although we have a (beautifully simple!) formula for this special case, we seek a general answer to count the number of spanning trees in an *arbitrary* graph. The answer to this is the so called Matrix Tree Theorem which provides a determinantal formula for the number of trees. As a result, we need to review some basics about linear algebra and determinants.

1.1 Determinants

Recall that a collection of distinct vectors $\vec{v}_1, \ldots, \vec{v}_s \in \mathbb{R}^n$ are said to be **linearly dependent** provided there exists scalars $a_1, \ldots, a_s \in \mathbb{R}$, not all zero, so that

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_s\vec{v}_s = \vec{0}.$$

Otherwise, we say they are **linearly independent**. For example, the fact that

$$2\begin{bmatrix} -7 \\ -1 \end{bmatrix} + \begin{bmatrix} 14 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

shows the vectors involved are linearly dependent. On the other hand, the vectors

$$\begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are linearly independent since there is no *non-zero* linear combination of them that gives the zero vector.

The next ingredient we needed is the idea of determinants. Recall that for a 2×2 matrix

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

we define the **determinant** of M as $\det M = ad - cb$. For example

$$\det\left(\begin{bmatrix} -7 & 14 \\ -1 & 2 \end{bmatrix}\right) = 0, \quad \det\left(\begin{bmatrix} 3 & 0 \\ 5 & 2 \end{bmatrix}\right) = 6, \qquad \det\left(\begin{bmatrix} 0 & 3 \\ 2 & 5 \end{bmatrix}\right) = -6$$

These examples further illustrate three important properties of determinants.

Properties of determinants:

1. If the rows (or columns) of M are linearly dependent, then $\det M = 0$.
2. If M is lower triangular, i.e., $b = 0$, then $\det M$ is the product of the numbers on its diagonal.
3. If M' is obtained by interchanging (or permuting) the rows or columns of M, then

$$\det M = \pm \det M'.$$

The reader should convince themselves of this for an arbitrary 2×2 matrix. More generally, one can define the determinant of any square matrix M. For example, in the 3×3 case we define

$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = a \cdot \det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b \cdot \det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c \cdot \det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right).$$

To help remember this formula, observe that the first term comes from multiplying a by the determinant of the 2×2 matrix obtained by deleting the row/column containing a. Similar constructions apply to the second and third term. The only twist, is that our sign alternates $+, -, +$ as we read across the top row. For example, we have

$$\det\left(\begin{bmatrix} 3 & 2 & 5 \\ 4 & 9 & 13 \\ 5 & 2 & 7 \end{bmatrix}\right) = 0, \quad \det\left(\begin{bmatrix} 3 & 0 & 0 \\ -1 & 4 & 0 \\ 7 & -1 & 5 \end{bmatrix}\right) = 60, \qquad \det\left(\begin{bmatrix} 0 & 3 & 0 \\ 0 & -1 & 4 \\ 5 & 7 & -1 \end{bmatrix}\right) = -60$$

Again, these examples illustrate the three aforementioned properties of determinants, in the 3×3 case. Although we do not prove that these properties hold for an arbitrary 3×3 matrix, one should not be totally surprised by this connection. After all, the 3×3 formula is intrinsically connected with the 2×2 case.

It now should come as little surprise that the general formula for a determinant is defined iteratively. Let

$$M = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}.$$

If $n = 1$, we define $\det M = a_{1,1}$. Otherwise, if $n > 1$,

$$\det M = a_{1,1} \det B_{1,1} - a_{1,2} \det B_{1,2} + a_{1,3} \det B_{1,3} - \cdots + (-1)^{n+1} a_{1,n} \det B_{1,n},$$

where $B_{i,j}$ is the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of M. You should convince yourself that this formula reduces to the formula given in the 2×2 and 3×3 cases.

A single computation by hand of the determinant of a 4×4 matrix is enough to convince one that this general formula is a bear! The good news is that we have little need for it. Instead, we only need the fact, which we do not prove here, that the general formula for determinants satisfy the three aforementioned properties.

Before closing this section we need one last result pertaining to determinants. Let A and B be $n \times m$ matrices $m \geq n$. If S is an n-element subset of the set $[m] = \{1, 2, \ldots, m\}$, then we define A_S to be the matrix obtained by deleting all the columns from A except those indexed by elements in S.

Theorem (Cauchy–Binet). *For A and B as above, we have*

$$\det(AB^t) = \sum_S \det(A_S)\det(B_S),$$

where our sum is over all n-element subsets of $[m] = \{1, 2, \ldots, m\}$ and t indicates transpose.

For example, if

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 0 & 5 \end{bmatrix} \qquad B^t = \begin{bmatrix} -1 & 0 \\ 3 & 4 \\ 7 & 2 \end{bmatrix},$$

then the Cauchy–Binet formula gives

$$\begin{aligned} \det(AB^t) &= \begin{vmatrix} 1 & 2 \\ 3 & 0 \end{vmatrix} \cdot \begin{vmatrix} -1 & 0 \\ 3 & 4 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ 3 & 5 \end{vmatrix} \cdot \begin{vmatrix} -1 & 0 \\ 7 & 2 \end{vmatrix} + \begin{vmatrix} 2 & 0 \\ 0 & 5 \end{vmatrix} \cdot \begin{vmatrix} 3 & 4 \\ 7 & 2 \end{vmatrix} \\ &= -6 \cdot -4 + 5 \cdot -2 + 10 \cdot -22 \\ &= -206. \end{aligned}$$

where the vertical bars indicate a determinant. Indeed, computing the left-hand side directly yields

$$\det(AB^t) = \det\left(\begin{bmatrix} 5 & 8 \\ 32 & 10 \end{bmatrix}\right) = -206.$$

1.2 Matrix Tree Theorem

In order to state the Matrix Tree Theorem we first recall that if G is a graph with vertices $v_1, \ldots, v_n$, then its **adjacency** matrix is the $n \times n$ matrix A whose (i, j)-entry is

$$a_{ij} = \begin{cases} 1 \text{ if } v_i \sim v_j \\ 0 \text{ otherwise} \end{cases}.$$

For our purposes we need a slight variation of the adjacency matrix called the **Laplacian of** G. It is defined as

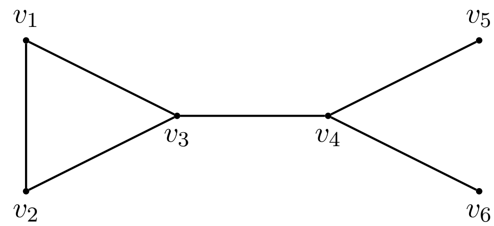
$$L = D - A$$

where

$$D = \begin{bmatrix} \deg v_1 & & & \\ & \deg v_2 & & \\ & & \ddots & \\ & & & \deg v_n \end{bmatrix}.$$

Lastly, we denote by L'' the matrix obtained by deleting the last row and last column from L.

For example if we consider the graph

G : 

then its Laplacian is

$$\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

Equipped with this new definition we now have the following famous theorem due to the physicist Gustav Kirchhoff (1824-1887).

Theorem 1 (Matrix Tree Theorem)**.** *Let G be a connected graph with Laplacian L. Then*

$$\det L'' = \tau(G).$$

Before proving this theorem (which takes some work) lets first use it to count the number of spanning trees in K_4. In this case,

$$L = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \qquad L'' = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

and, indeed $\det L'' = 3 \cdot 8 - 4 - 1 \cdot 4 = 16 = 4^2$. So at least in this case, our new formula coincides with Cayley's formula. This suggests another proof of Cayley's formula via the Matrix Tree Theorem. Indeed,

$$L''_{K_n} = \underbrace{\begin{bmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{bmatrix}}_{n-1}.$$

If we now subtract the bottom row from each of the other rows we get

$$A = \left[\begin{array}{cccc|c} n & & 0 & & -n \\ & n & & & -n \\ 0 & & \ddots & & \vdots \\ & & & n & -n \\ \hline -1 & -1 & \dots & -1 & n-1 \end{array}\right],$$

If we now add each of the first $n-2$ columns to the last column we get

$$B = \left[\begin{array}{cccc|c} n & & 0 & & 0 \\ & n & & & 0 \\ 0 & & \ddots & & \vdots \\ & & & n & 0 \\ \hline -1 & -1 & \dots & -1 & 1 \end{array}\right].$$

As the operation of adding a row/column in a matrix to another row/column preserves determinants we obtain:

$$\tau(K_n) = \det(L''_{K_k}) = \det A = \det B = n^{n-2},$$

where the last equality follows as B is lower triangular with $n-2$ copies of n and one 1 on its diagonal. This completes our (second) proof of Cayley's formula.

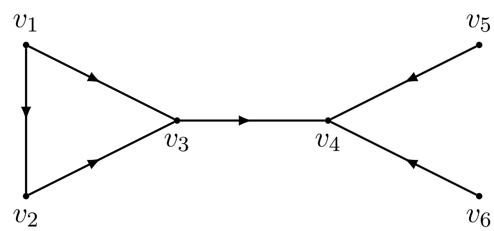
We now turn our attention to the proof of the Matrix Tree Formula. Instead of diving into the details right away, we structure our argument as follows. Using an example, we motive two key lemmas. Using these lemmas we then prove the main result. Lastly we prove these lemmas.

To motivate our first step, recall that the formula for a determinant involves an alternating plus/minus component. Interpreting sign as a positive/negative direction, our first definition feel less like a "leap" and more like a "step".

Definition. A **directed graph** is graph G where each edge is assigned a direction or orientation. We indicate this direction by drawing an arrow along each edge e and define the **head** of e to be the vertex head(e) our arrow points toward and the **tail** of e to be the vertex tail(e) our arrow points from.

For example, in the directed graph

G :



we have head(v_4v_5) $= v_4$ and tail(v_4v_5) $= v_5$.

Although we cannot encode a directed graph with either an adjacency or Laplacian matrix, we can encode it with something called an *incident matrix*.

Definitions. Let G be a directed graph with vertices $v_1, \ldots, v_n$ and directed edges $e_1, \ldots, e_m$. We define its **incident matrix** to be the $n \times m$ matrix $M = (m_{ij})$ where

$$m_{ij} = \begin{cases} 1 & \text{if head}(e_j) = v_i \\ -1 & \text{if tail}(e_j) = v_i \\ 0 & \text{otherwise} \end{cases}.$$

As the rows and columns of M are indexed by the vertices and edges, respectively, of G, we denote m_{ij} by $M(v_i, e_j)$. Furthermore, we denote the jth column of M by $M(e_j)$. Lastly, let M' be the matrix obtained by deleting the last row of M.

For example, if we enumerate the edges in the direct graph G above as

$$e_1 = v_1v_3,\ e_2 = v_1v_2,\ e_3 = v_2v_3,\ e_4 = v_3v_4,\ e_5 = v_4v_5,\ \text{ and } e_6 = v_4v_6$$

then its incident matrix is

$$M = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Lets looks a bit closer at this example. In particular, consider the edges e_1, e_2, and e_3 which form a cycle in G. If we traverse this cycle in a counterclockwise fashion starting at v_1, then we traverse e_2 and e_3 in the direction of their orientation but traverse e_3 opposite to its orientation. Notice that if we consider the following linear combination

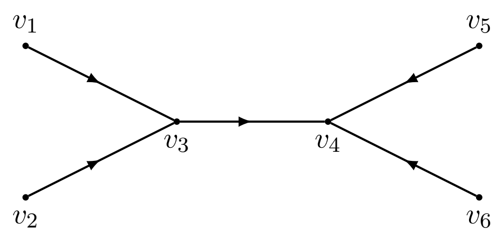
$$M(e_2) + M(e_3) - M(e_1),$$

where the -1 corresponds to the fact that e_1 is walked opposite its orientation, we see that it simplifies to $\vec{0}$. In this case, this shows that the columns of M are linearly dependent. This motivates our first lemma, which states that any graph with at least one cycle must have an incident matrix with linearly dependent columns.

Lemma 1. *Let G be a directed graph with at least one cycle. Then the columns of its incident matrix M are linearly dependent. In particular, if G has order n and size $n-1$, then* $\det M' = 0$.

What if our graph contains no cycles? In this case what can be said about its incident matrix? To motivate our answer, consider the subgraph T of our running example G

T :



Now imagine pulling leaves off our tree in some manner until we are left with v_6. One way to do this is

$$v_1, v_2, v_3, v_5, v_4, v_6$$

Doing so deletes (in order) the edges $13, 23, 34, 45, 46$. With this ordering of vertices/edges we have

$$M_T = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

and therefore the matrix obtained from M_T by deleting its last row, M_T', is lower triangular with ± 1 along its diagonal. Therefore $\det(M_T') = \pm 1$. This gives us our next lemma.

Lemma 2. *Let T be a tree with vertices $v_1, \ldots, v_n$ and edges $e_1, \ldots, e_m$ and let M be the resulting incident matrix. Then the rows and columns of M' can be reordered to give an lower triangular matrix with ± 1's along its diagonal. Therefore* $\det M' = \pm 1$.

Lemma 3. *Let G be a graph. For some ordering of its vertices and edges, let L and M be its Laplacian and incident matrix, respectively. Then*

$$L'' = M'(M')^t$$

where $(M')^t$ is the transpose of M'.

Proof of the Matrix Tree Theorem. Assume G has order n and size m. For some ordering of its vertices and edges, let L and M be its Laplacian and incident matrix, respectively. Furthermore, set L'' and M' to be the matrix obtained by deleting the last row of L and M respectively. We now have the following equalities,

$$\det L'' = \det\left(M'(M')^t\right) = \sum_{\substack{S\subseteq[m]\\|S|=n-1}} \det M'_S \det M'_S = \sum_{\substack{H\subseteq G\\ \text{order } n\\ \text{size } n-1}} \left(\det M'_T\right)^2 = \sum_{\substack{T \text{ span. tree}\\ \text{of } G}} \left(\det M'_H\right)^2 = \tau(G),$$

where the first equality follows from Lemma 3, the second from Cauchy-Binet, the fourth is Lemma 1 and the fact that any acyclic subgraph of G with order n and size $n-1$ is a tree and hence a spanning tree. The final equality follows from Lemma 2.

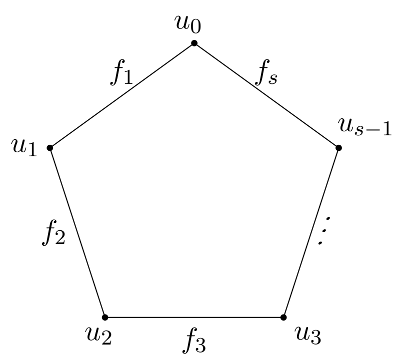
□

With the proof of The Matrix Tree theorem complete, we now turn our attention to proving Lemmas 1,2, and 3.

Proof of Lemma 1. Let $C = (u_0, u_1, \ldots, u_s = u_0)$ be a cycle in G and define its edges to be

$$f_k = u_{k-1}u_k$$

for $1 \le k \le s$. In other words we have



so that if we traverse C starting at u_0 we first traverse f_1, then f_2, etc. Doing so, we may traverse an edge opposite to its orientation. (Think of walking the wrong way down a one-way street!) To quantify this, define the sequence $d_1, d_2, \ldots, d_s$ so that

$$d_k = \begin{cases} 1 & \text{if } \operatorname{head}(f_k) = u_k \\ -1 & \text{if } \operatorname{tail}(f_k) = u_k \end{cases}$$

We now claim

$$d_1M(f_1) + d_2M(f_2) + \cdots + d_sM(f_s) = \vec{0}. \qquad (\star)$$

In other words, the column vectors in M corresponding to the edges in C are linearly dependent as needed.

To justify $(\star)$ consider a vertex v. If v is not on C, then $M(v, f_k) = 0$ for all k. Next consider a vertex u_k on C. As f_k is incident to u_k, then

$$d_k = M(u_k, f_k),$$

so $d_kM(u_k, f_k) = 1$. Furthermore, since

$$\operatorname{head}(f_{k+1}) = u_{k+1} \Longleftrightarrow \operatorname{tail}(f_{k+1}) = u_k$$

and

$$\text{tail}(f_{k+1}) = u_{k+1} \Longleftrightarrow \text{head}(f_{k+1}) = u_k$$

it follows that

$$d_{k+1} = -M(u_k, f_{k+1}).$$

and hence $d_{k+1}M(u_k, f_{k+1}) = -1$. Finally, since $M(u_k, e) = 0$ for any edge e in C such that $e \neq f_k, f_{k+1}$, we see that $(\star)$ holds. □

Proof of Lemma 2. Following our motivation for the statement of this lemma, we imagine removing leafs from our tree T until we are left with the vertex v_n. Ordering our vertices according to such a sequence of deletions gives the sequence $u_1, \ldots, u_n$ where $u_n = v_n$. This deletion sequence also gives rise to a the edge ordering $f_1, \ldots, f_m$ where f_1 is the first edge deleted, f_2 the second, etc. We now observe that u_i is only incident to edges f_j where $j \leq i$. Moreover u_i is incident f_i. If N is the incident matrix for T under this vertex/edge ordering, then this observation directly implies that N is lower triangular with ± 1's on its diagonal. Hence

$$\det M' = \det N' = \pm 1.$$

□

Proof of Lemma 3. Let $v_1, \ldots, v_n$ and $e_1, \ldots, e_m$ be the vertices and edges of G. To show this, we first consider the diagonal elements of the Laplacian. For any $i < n$ we know that the (i,i)-entry of L'' is $\deg v_i$. Indeed since the number of ± 1's on the ith row of M is just the degree of v_i, then the (i,i)-entry of $M'(M')^t$ is also $\deg v_i$.

Now consider when $i \neq j$. In this case the (i,j)-entry of L'' is -1 iff $v_i \sim v_j$. On the other side of our equation, the (i,j)-entry of $M'(M')^t$ is the dot product of the ith and jth row of M'. From the definition of the incident matrix, it follows that there is at most one column that contains two nonzero entries in these two rows. Moreover, this occurs with a 1 in one of the rows and a -1 in the other iff $v_i \sim v_j$. Hence the (i,j)-entry of $M'(M')^t$ is -1 iff $v_i \sim v_j$. This completes the proof. □