An easy bijective proof of the Matrix-Forest Theorem

Kimmo Eriksson
Dept. of Mathematics, SU, S-106 09 Stockholm, Sweden
kimmo@nada.kth.se

Abstract

The Matrix-Forest Theorem says that for a subset \( I \) of vertices of a digraph, the number of \( I \)-rooted spanning forests is the determinant of the submatrix obtained from the Laplacian by deleting all rows and columns corresponding to nodes in \( I \). We give an easy bijective proof of this fact.

The rather well-known Matrix-Tree Theorem gives the number of spanning trees of a graph as a minor of the Laplacian of the graph. This note will give an easy bijective proof of what one might call the "Matrix-Forest Theorem", which is a slightly less general version of the "All Minors Matrix-Tree Theorem" of Chen [2] and Chaiken [3], while containing the ordinary Matrix-Tree Theorem, see Biggs [1] or Goulden-Jackson [4]. This new proof should be good for education purposes, taking the theorem down to a bijective interpretation of the expansion of the determinant.

Let \( G = (V, E) \) be a finite directed graph. Between any pair of nodes there may be arbitrarily many edges, and they are distinguishable. Let \( d_{ij} \) denote the number of edges directed from node \( i \) to node \( j \) in \( G \). Let \( d_{i}^+ = \sum_{j \neq i} d_{ij} \), the total number of edges that are directed from \( i \) to any other node; loops are disregarded. The Laplacian of \( G \) is the square matrix \( L \in \mathbb{R}^{V \times V} \) where the element in place \((i, j)\) is \( d_{ij}^+ \) if \( i = j \) and \(-d_{ij}\) if \( i \neq j \). The row sums of \( L \) are zero, so \( L \) is singular.

A \( J \)-cycle in \( G \), for \( J \subseteq V \), is a directed cycle visiting each of the nodes in \( J \) once.

For any node \( i \), by an \( i \)-rooted tree in \( G \) we mean a tree where every node has exactly one outgoing edge except \( i \) which has none. In other words, for any node \( j \) in the tree, the path between \( j \) and \( i \) in the tree is directed towards \( i \). (Such a tree is sometimes called an "in-directed arborescence".) If the tree reaches all the nodes in \( V \), it is an \( i \)-rooted spanning tree.

For any node set \( I \subseteq V \), by an \( I \)-rooted spanning forest in \( G \) we mean a collection of \( i \)-rooted trees, one for each \( i \in I \), such that every node in \( V \) is in exactly one of

the components. For example, an \( \{i\}\)-rooted spanning forest is simply an \( i \)-rooted spanning tree; there exists no \( \emptyset \)-rooted spanning forest.

Theorem 1 (Matrix-Forest Theorem) If \( G = (V, E) \) is a digraph with Laplacian \( L \) and \( I \subseteq V \), then the number of \( I \)-rooted spanning forests in \( G \) is

\[
\text{spanf}_G(I) = \det L_I,
\]

where \( L_I \) denotes the submatrix obtained by deleting the \( i \)-th row and column for all \( i \in I \).

The theorem will follow from this lemma:

Lemma 1 \( \text{spanf}_G(I) \) can be determined recursively by

\[
\text{spanf}_G(I) = d_x^+ \cdot \text{spanf}_G(I \cup \{x\}) - \sum_{x \in J \subseteq V \setminus I \atop |J| \geq 2} \# J\text{-cycles} \cdot \text{spanf}_G(I \cup J),
\]

where \( x \) is any node in \( V \setminus I \).

Proof. The first term on the righthand side counts pairs consisting of one edge directed from \( x \) and one \( (I \cup \{x\}) \)-rooted spanning forest. When adding an edge \( e \) directed from \( x \) to some node \( y \) to a \( (I \cup \{x\}) \)-rooted spanning forest, we get one of two cases, depending on which of the trees in the forest that contains \( y \):

Case 1: \( y \) is a node in a tree rooted in \( i \in I \). Then by adding \( e \) the \( x \)-rooted tree becomes a part of the \( i \)-rooted tree, and what remains is an \( I \)-rooted spanning forest.

Case 2: \( y \) is a node in the \( x \)-rooted tree. Then by adding \( e \) we get some directed cycle \( C = (xy \cdots) \). Say that \( J \) is the set of nodes in \( C \), so \( C \) is a \( J \)-cycle. Erase the edges in this cycle. Then all the nodes in \( J \) become roots, so we obtain an \( I \cup J \)-rooted spanning forest.

It is easily seen that the steps above are invertible. Hence we have described a bijection that establishes the identity

\[
d_x^+ \cdot \text{spanf}_G(I \cup \{x\}) = \text{spanf}_G(I) + \sum_{x \in J \subseteq V \setminus I \atop |J| \geq 2} \# J\text{-cycles} \cdot \text{spanf}_G(I \cup J),
\]

which is equivalent to the desired formula. □

We shall now relate the number of \( I \)-rooted spanning forests of \( G \) to determinants of certain submatrices of the Laplacian \( G \). Let us adopt the following conventions: The determinant of an empty submatrix is 1; \( S_V \) is the set of permutations on the set \( V \); \( C_V \subset S_V \) is the set of cyclic permutations on \( V \).

We shall prove the theorem by expanding the determinant in the cycles containing a certain element. An alternative form of the basic expansion of the determinant of a matrix \( A \in \mathbb{R}^{V \times V} \) with \( x \in V \) is

\[
\det A = a_{xx} \det A_{\{x\}} - \sum_{x \in J \subseteq V} \sum_{x \in C_J} \prod_{j \in J} (-a_{i,j}) \det A_J.
\]  \hspace{1cm} (1)

302
This can be obtained as follows from the basic expansion,
\[ \det A = \sum_{\pi \in \mathcal{S}_V} \text{sgn } \pi \prod_{j \in V} a_{j, \pi(j)}. \]

Split the sum into two parts according to whether \( x \) is a fixpoint of \( \pi \) or not. If \( x \) is not a fixpoint, let \( \tau \) be the cycle that contains \( x \) in the cycle decomposition of \( \pi \), and let \( J \) be the set of elements in the cycle. Thus we have \( x \in J, \tau \in \mathcal{C}_J \) and, because \( x \) was not a fixpoint, \( |J| \geq 2 \).

\[ \det A = a_{xx} \sum_{\pi' \in \mathcal{S}_{V \setminus \{x\}}} \text{sgn } \pi' \prod_{j \in V \setminus \{x\}} a_{j, \pi'(j)} + \sum_{x \in J \subseteq V} \sum_{\tau \in \mathcal{C}_J, |J| \geq 2} \text{sgn } \tau \prod_{j \in J} a_{j, \tau(j)} \sum_{\sigma \in \mathcal{S}_{V \setminus J}} \text{sgn } \sigma \prod_{k \in V \setminus J} a_{k, \sigma(k)}. \]

Since \( \tau \) is a cycle on \( J \), the sign of \( \tau \) is \( \text{sgn } \tau = (-1)^{|J|+1} \). Hence we can multiply each \( a_{j, \tau(j)} \) by a \((-1)\) and still have one \((-1)\) left. By using the basic expansion of the determinant again, twice, we get equation (1).

The theorem is now proved by induction on \( |V \setminus I| \), the size of submatrix \( L_I \). If \( |V \setminus I| = 0 \), i.e. if \( V = I \), then \( L_I \) is the empty matrix. Since there is only one \( V \)-rooted spanning forest and \( \det L_I = 1 \), the statement is true. Suppose it is true whenever \( |V \setminus I| \leq p \) and consider a subset \( I \subseteq V \) with \( |V \setminus I| = p + 1 > 0 \). Choose some \( x \in V \setminus I \). Equation 1, with \( A = L_I, a_{xx} = d^+_x \) and \(-a_{ij} = d_{ij} \) when \( i \neq j \), gives:

\[ \det L_I = d^+_x \cdot \det L_{I \cup \{x\}} - \sum_{x \in J \subseteq V \setminus I} \sum_{\tau \in \mathcal{C}_J, |J| \geq 2} \prod_{j \in J} d_{j, \tau(j)} \det L_{I \cup J}. \]

Now, thanks to the induction hypothesis, \( \det L_{I \cup \{x\}} \) is equal to the number of \( \{I \cup \{x\}\}\)-rooted spanning forests in \( G \); also, \( \det L_{I \cup J} \) is the number of \( \{I \cup J\}\)-rooted spanning forests. The sum \( \sum_{\tau \in \mathcal{C}_J} \prod_{j \in J} d_{j, \tau(j)} \) is clearly the number of \( J \)-cycles in \( G \). Hence we have proved that

\[ \det L_I = d^+_x \cdot \text{span}_{G}(I \cup \{x\}) - \sum_{x \in J \subseteq V \setminus I} |J| \text{-cycles} \cdot \text{span}_{G}(I \cup J) = \text{span}_{G}(I) \]

by the lemma. The theorem follows by induction. \( \square \)

References


(Received 11/4/95)