A simple proof of the matrix-tree theorem, upward routes, and a matrix-tree-cycle theorem

Jules Jacobs

July 25, 2021

Abstract

The matrix-tree theorem states that the number of spanning trees in a graph is $\det L'$, where L' is the Laplacian matrix L of the graph, with any row and column deleted. We give a direct proof of the fact that $\det(xI+L)$ is the generating function of spanning forests with k roots. Our proof does not rely on the Cauchy-Binet formula, yet is arguably simpler than the standard proof, and applies to weighted & directed graphs as well.

The proof is based on a lemma that if A is the adjacency matrix of a graph where each vertex has at most one outgoing edge, then $\det(I-A)=1$ if A is a forest and $\det(I-A)=0$ otherwise. We generalize this lemma to any graph, in which case $\det(I-xA)^{-1}$ is shown to be the generating function of *upward routes*.

Lastly, we generalize the matrix-tree theorem to a theorem about $\det(A+L)$ where A is the adjacency matrix of a second graph, which reduces to counting spanning forests when $A=xI$ but allows some cycles when A is not diagonal. The special case $L=0$ gives that $\det(A)$ counts signed cycle covers. The all-minors matrix-tree theorem follows as a corollary. For instance, the fact that $\det(L')$ counts spanning forests, where L' is L with any column i and row j deleted, follows by picking A to have $A_{ij}=1$ and zero elsewhere.

1 Introduction

Given finite sets of numbers $T_1,\ldots,T_n \subset \mathbb{R}$, we have the identity:

$$\left(\sum_{x_1\in T_1} x_1\right)\cdots\left(\sum_{x_n\in T_n} x_n\right) = \sum_{x_1\in T_1}\cdots\sum_{x_n\in T_n} x_1\cdots x_n$$

On the right hand side, we get one term for every way of choosing $(x_1,\ldots,x_n)\in T_1\times\cdots\times T_n$. Similarly, given finite sets of vectors $T_1,\ldots,T_n\subset\mathbb{R}^k$, we have the following identity, by multilinearity of the determinant:

$$\det\left(\sum_{v_1\in T_1} v_1 \,\middle|\, \cdots \,\middle|\, \sum_{v_n\in T_n} v_n\right) = \sum_{v_1\in T_1}\cdots\sum_{v_n\in T_n}\det(v_1|\cdots|v_n)$$

Where $\det(v_1|\ldots|v_n)$ is the determinant of a matrix with those columns. We shall see that Kirchoff's theorem can be obtained from this identity, but we first need some definitions and a lemma.

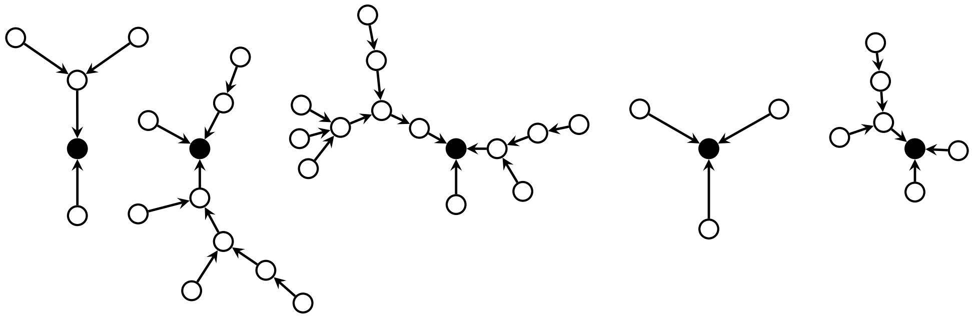

Figure 1: A forest with 5 roots.

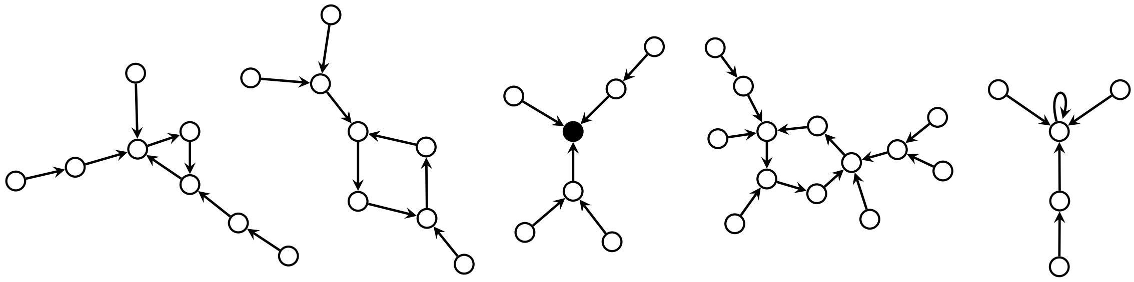

Figure 2: A 01-graph with one root. If we delete the third component, we obtain a 1-graph.

Definition 1.1. We define the concepts *1-graph*, *01-graph*, *forest*, *root*, and *tree*:

- A *1-graph* is a directed graph where each vertex has 1 outgoing edge.
- A *01-graph* is a directed graph where each vertex has 0 or 1 outgoing edges.
- A *forest* is a 01-graph with no cycles.
- The *roots* are vertices with 0 outgoing edges.
- A *tree* is a forest with one root.

See Figure 1 and Figure 2 for examples.

If we start at a vertex v in a 01-graph and keep following the unique outgoing edge, then we either loop in a cycle, or we end up at a vertex with no outgoing edges. Therefore a general 01-graph consists of roots and cycles, plus trees converging onto the roots and cycles. A 1-graph is a 01-graph with no roots, so regardless of where we start, we always end up in a cycle.

2 The matrix-tree theorem

We need a lemma that give us an indicator function for forests.

Lemma 2.1. *Let A_G be the adjacency matrix of a 1-graph G, then*

$$\det(I - A_G) = \begin{cases} 1 & \text{if } G \text{ is empty} \\ 0 & \text{if } G \text{ is not empty} \end{cases}$$

Proof. If G is empty, we have a 0×0 matrix, which has determinant 1. If G is not empty, then each column of $I - A_G$ has one $+1$ diagonal entry and one -1 entry from A_G, so the sum of the rows is zero, so $\det(I - A_G) = 0$. This remains true in the presence of self loops. □

Lemma 2.2. *Let A_G be the adjacency matrix of a 01-graph G, then*

$$\det(I - A_G) = \begin{cases} 1 & \text{if } G \text{ is a forest} \\ 0 & \text{if } G \text{ has a cycle} \end{cases}$$

Proof. We calculate the determinant by repeatedly performing Laplace expansion on a column i that corresponds to a root. The column of a root has a single $+1$ entry on the diagonal, so performing Laplace expansion along this column deletes column i and row i. Row i contains all the incoming edges of the root. Therefore, this operation corresponds to deleting root i and all its incoming edges from the graph. Deleting a root may create new roots. Repeating this process of deleting roots, the remaining 1-graph will be empty iff the original graph was a forest. Applying the previous lemma gives the desired result. □

An alternative shorter proof using eigenvalues:

Proof. If G is a forest then A_G is nilpotent, so all its eigenvalues are 0, so all the eigenvalues of $I - A_G$ are 1, so $\det(I - A_G) = 1$. If G has a cycle, then A_G has 1 as an eigenvalue (take the eigenvector that is 1 on the cycle and 0 elsewhere), so $I - A_G$ has 0 as an eigenvalue, so $\det(I - A_G) = 0$. □

We now fix $n \times n$ matrices A and D over some commutative ring:

- An arbitrary matrix A of edge weights (with A_{ij} being the weight of edge $i \to j$)
- A diagonal matrix D of vertex weights (with D_{ii} being the weight of vertex i).

Definition 2.1. The Laplacian matrix L is defined as having columns:

$$L_i = \sum_j A_{ij}(e_i - e_j)$$

Definition 2.2. The weight of a forest G is:

$$w(G) = \prod_{i \in \text{roots}(G)} D_{ii} \prod_{(i \to j) \in \text{edges}(G)} A_{ij}$$

We're ready to state Kirchoff's theorem for multiple-root forests in weighted directed graphs.

Theorem 2.1 (Kirchoff, Tutte)**.** *The determinant* $\det(D + L)$ *is the weight-sum of all forests on n vertices:*

$$\det(D + L) = \sum_{\textit{forest } G} w(G)$$

Proof. The i-th column of the matrix is $(D + L)_i = D_{ii}e_i + \sum_j A_{ij}(e_i - e_j)$, so

$$\det(D+L) = \det \left(\begin{array}{c|c|c|c} D_{11}e_1 & D_{22}e_2 & & D_{nn}e_n \\ + & + & & + \\ A_{11}(e_1 - e_1) & A_{21}(e_2 - e_1) & & A_{n1}(e_n - e_1) \\ + & + & \cdots & + \\ A_{12}(e_1 - e_2) & A_{22}(e_2 - e_2) & & A_{n2}(e_n - e_2) \\ \vdots & \vdots & & \vdots \\ A_{1n}(e_1 - e_n) & A_{2n}(e_2 - e_n) & & A_{nn}(e_n - e_n) \end{array} \right) = \sum_{\text{01-graph } G} w(G) \det(I - A_G) = \sum_{\text{forest } G} w(G)$$

The first step is by the definition of L. In the middle step we expand the determinant by multilinearity: we sum over all possible ways to pick one term of each column. To choose a 01-graph on n vertices, is to choose for each vertex i (column i) whether to make i a root (term $D_{ii}e_i$) or to give i an outgoing edge $i \to j$ (term $A_{ij}(e_i - e_j)$). Then we take the weights D_{ii} and A_{ij} out of the determinant, and we're left with $w(G) \det(I - A_G)$, where A_G is the adjacency matrix of the chosen 01-graph. The final step is applying lemma 2.2. □

The classic form of Kirchoff's matrix tree theorem gives us a way to count the number of spanning trees of an *undirected* and *unweighted* graph G. It is a special case of Theorem 2.1, as follows.

We interpret the undirected unweighted graph as a directed weighted graph where we insert two edges $i \longrightarrow j$ and $j \longrightarrow i$ of weight 1 for each undirected edge i — j. This makes the matrix A the adjacency matrix of G:

$$A_{ij} = \begin{cases} 1 & \text{if } \{i,j\} \in G \\ 0 & \text{otherwise} \end{cases}$$

The Laplacian L is still given by definition 2.1.

Corollary 2.1.1 (Classic form of Kirchoff's theorem)**.** Let L' be the Laplician matrix L with the first row and column deleted. Then $\det(L')$ is the number of spanning trees of G.

Proof. Set the vertex weight $D_{11} = 1$ for vertex 1 and $D_{ii} = 0$ for the other vertices. Then on the one hand, $\det(D + L) = \det(L')$, and on the other hand, $\det(D + L)$ is the number of directed trees with 1 as the root, by theorem 2.1. Such trees are in bijective correspondence with undirected spanning trees of G, because we can turn a tree rooted at 1 into an undirected spanning tree by forgetting the direction of the edges, and we can turn an undirected spanning tree into a tree rooted at 1 by directing all the edges toward 1. □

More generally, theorem 2.1 allows us to count the number of forests with a given number of roots. For our commutative ring we pick polynomials in x and set $D = xI$. Theorem 2.1 gives that the (slightly modified) characteristic polynomial $\det(xI + A)$ of the adjacency matrix A is the generating polynomial of forests with k roots.

It is essential that we consider directed forests. The correspondence between undirected spanning trees and directed spanning trees rooted at 1 fails to work as smoothly for $k > 1$. Thus it could be argued that Kirchoff's theorem is really a theorem about directed forests. The directed version was Tutte's contribution to the theorem.

3 Upward routes

We now know that when A is the adjacency matrix of a 01-graph, then $\det(I - A) = 1$ if G is a forest and $\det(I - A) = 0$ if G has a cycle. One naturally wonders about the value of $\det(I - A)$ when A is an arbitrary adjacency matrix. The goal of this section is a combinatorial interpretation of $\det(I - xA)$ and $\det(I - xA)^{-1}$ for an arbitrary adjacency matrix, which generalizes the lemma to arbitrary graphs.

Definition 3.1. Given a directed graph G with an order on the vertices, we define *(strictly) upward loops* and *(strictly) upward routes*:

- An *upward loop* at vertex i is a walk from i to i that does not visit vertices lower than i.
- A *strictly upward loop* at vertex i is a walk from i to i that only visits vertices higher than i (except at the start/endpoint of the walk, where it does visit i itself).
- A *(strictly) upward route* is a choice of (strictly) upward loop at each vertex.

Let $f_i(x)$ be the generating function of strictly upward loops of length k at vertex i. Then

$$f_i^*(x) = (1 - f_i(x))^{-1}$$

is the generating function of upward loops of length k at vertex i, because an upward loop splits uniquely into a sequence of strictly upward loops. Furthermore, the generating functions $f(x)$ and $f^*(x)$ of (strictly) upward routes of k edges are given by:

$$f(x) = \prod_{i=1}^{n} f_i(x) \qquad\qquad f^*(x) = \prod_{i=1}^{n} f_i^*(x)$$

Recall Cramer's rule:

Theorem 3.1. *(Cramer's rule) Let A be an invertible matrix and let $A_{[i,j]}$ be the matrix A with column i and row j deleted, then:*

$$(A^{-1})_{ij} = \frac{\det(A_{[i,j]})}{\det(A)}$$

From this, we get the following lemma that allows us to calculate $\det(A)^{-1}$ in terms of entries of inverses of submatrices of A:

Lemma 3.1. *Given a matrix A, provided both sides are defined,*

$$\det(A)^{-1} = \prod_{i=0}^{n-1} (A_{[1\ldots i,1\ldots i]})^{-1}_{11}$$

Where $A_{[1\ldots i,1\ldots i]}$ is the matrix A with the first i rows and columns deleted.

Proof. Cramer's rule implies:

$$\det(A)^{-1} = A^{-1}_{1,1} \cdot \det(A_{[1,1]})^{-1}$$

Continuing this by induction, we get:

$$\begin{aligned}\det(A)^{-1} &= A^{-1}_{1,1} \cdot \det(A_{[1,1]})^{-1} \\ &= A^{-1}_{1,1} \cdot (A_{[1,1]})^{-1}_{11} \cdot \det(A_{[1..2,1..2]})^{-1} \\ &= \cdots \\ &= A^{-1}_{1,1} \cdot (A_{[1,1]})^{-1}_{11} \cdot (A_{[1..2,1..2]})^{-1}_{1,1} \cdots (A_{[1..n-1,1..n-1]})^{-1}_{1,1} \cdot 1\end{aligned}$$

□

Lemma 3.2. *The generating function of upward loops at vertex i is $((I - xA)_{[1\ldots i-1,1\ldots i-1]})^{-1}_{11}$.*

Proof. Given an adjacency matrix B, $(I - xB)^{-1} = I + xB + x^2B^2 + \cdots$ is the matrix generating function of walks, so $(I - xB)^{-1}_{11}$ is the generating function of loops from vertex 1 to 1. To obtain upward loops at vertex i in A we take $B = (I - xA)_{[1\ldots i-1,1\ldots i-1]}$ with the first $i - 1$ rows and columns deleted. □

We combine these lemmas to obtain:

Theorem 3.2. *The generating function of upward routes with k edges is* $\det(I - xA)^{-1}$.

Proof. Combine the preceding two lemmas with $f^*(x) = \prod_{i=1}^{n} f_i^*(x)$. □

Corollary 3.2.1. The number of upward routes of k edges does not depend on the order of the vertices.

Proof. If we permute the order of the vertices by a permutation P, the generating function

$$\det(I - xPAP^{-1}) = \det(P(I - xA)P^{-1}) = \det(I - xA)$$

stays the same. A bijective proof is left as an exercise. □

Corollary 3.2.2. For an arbitrary adjacency matrix A,

$$\det(I - xA) = \prod_{i=1}^{n} (1 - f_i(x))$$

Where $f_i(x)$ is the generating function of strictly upward loops at vertex i.

Proof. Use the relationship between the generating functions:

$$\det(I - xA) = f^*(x)^{-1} = \prod_{i=1}^{n} f_i^*(x)^{-1} = \prod_{i=1}^{n} (1 - f_i(x))$$

□

Note that $\det(I - xA)$ is a polynomial, even though $f_i(x)$ is a power series. If we take the coefficient of x^k of Corollary 3.2.2, then modulo sign conventions we obtain precisely Theorems 1&2 from [Rot01] about clow sequences (clow sequences are equivalent to strictly upward routes). Corollary 3.2.2 is equivalent to Theorem 2 from [Rot01], but is stated directly in terms of the polynomials, rather than in terms of their coefficients.

The main lemma follows as a corollary, which gives us a third proof of the lemma:

Corollary 3.2.3. Let A_G be the adjacency matrix of a 01-graph G, then

$$\det(I - A_G) = \begin{cases} 1 & \text{if } G \text{ is a forest} \\ 0 & \text{if } G \text{ has a cycle} \end{cases}$$

Proof. If G is a forest, then it has no strictly upward loops, so $f_i(x) = 0$, so $\det(I - xA_G) = 1$. If G has a cycle, let i be the lowest vertex on the cycle. Then $f_i(x) = x^k$ where k is the length of the cycle. Now substitute $x = 1$ to obtain $\det(I - A_G) = 0$. □

4 The matrix-tree-cycle theorem

We shall now generalize the matrix-tree theorem about $\det(D + L)$ to the *matrix-tree-cycle theorem* by allowing D to be a general matrix. This theorem will express $\det(D + L)$ as a sum over 1-graphs with each edge labeled either with T (for tree) or with C (for cycle) but not both.

Definition 4.1. A TC-labeled 1-graph is called *tree-cyclic* if:

1. Each cycle has at least one C-edge (*i.e.,* the T-edges form a forest)
2. Each C-edge lies on a cycle (*i.e.,* edges not part of a cycle must be T-edges)

The cycle that a C-edge is a part of may use other C-edges as well as T-edges, but a cycle cannot consist solely of T-edges. Said differently: an 1-graph consists of a set of cycles and some extra edges converging onto those cycles. We obtain a valid tree-cyclic TC labeling as long as we obey two constraints: we must label at least one edge on each cycle with C, and we must label all the converging edges with T. The remaining cycle edges can be labeled with either C or T. Figure 3 contains an example of a tree-cyclic graph.

Definition 4.2. The sign of a TC-labeled 1-graph G is:

$$(-1)^G = (-1)^{\#\text{cycles}+\#\text{C-edges}}$$

where #cycles is the number of cycles of G and #C-edges is the total number of C-labeled edges in G.

We associate a matrix M_G with a TC-labeled 1-graph G. This matrix will play the role that $I - A_G$ played in the matrix-tree theorem.

Definition 4.3. The matrix M_G is defined as having columns:

$$(M_G)_i = \begin{cases} e_j & \text{if the outgoing edge } i \to j \text{ has label C} \\ e_i - e_j & \text{if the outgoing edge } i \to j \text{ has label T} \end{cases}$$

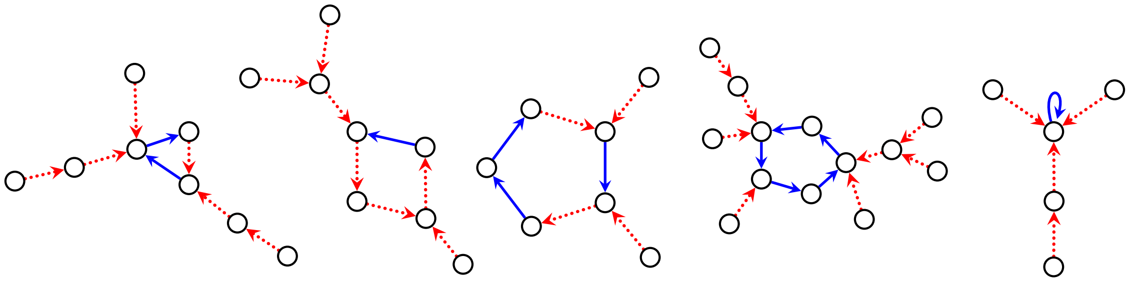

Figure 3: A tree-cyclic 1-graph. T-edges are red dotted edges, and C-edges are solid blue edges.

The following lemma generalizes the main lemma of the matrix-tree theorem.

Lemma 4.1. *The determinant of M_G is given by:*

$$\det(M_G) = \begin{cases} (-1)^G & \text{if } G \text{ is tree-cyclic} \\ 0 & \text{otherwise} \end{cases}$$

Proof. Calculate the determinant in steps:

1. Perform Laplace expansion on vertices with no predecessors. This makes the determinant zero if the successor is a C-edge, and deletes the corresponding vertex if the successor is a T-edge.
2. We are now left with a disjoint set of cycles. If there is a cycle consisting solely of T-edges, the determinant is zero because the corresponding rows sum to zero.
3. We are now left with a disjoint set of cycles where each cycle has at least one C-edge. Use the C-edges to turn the T-edges into C-edges by row operations. The determinant obtains a -1 sign for each such switch.
4. We are now left with a disjoint set of cycles where each cycle consists solely of C-edges. In other words, a permutation matrix. The determinant of a permutation matrix is $(-1)^{\#\text{cycles}+\#\text{edges}}$

If there was a C-edge not part of a cycle we have obtained 0 in step 1, and if there was a cycle among T-edges we have obtained 0 in step 2. For the remaining graphs, we have obtained a -1 sign for each T-edge in the cycles, so together with $(-1)^{\#\text{cycles}+\#\text{edges}}$ we are left with $(-1)^{\#\text{cycles}+\#\text{C-edges}} = (-1)^G$. □

We now fix $n \times n$ matrices A and D over some ring:

- An arbitrary matrix A of weights for T-edges (with A_{ij} being the weight of edge $i \xrightarrow{T} j$)
- An arbitrary matrix D of weights for C-edges (with D_{ij} being the weight of edge $i \xrightarrow{C} j$).
- As before, the Laplacian L is given by $L_i = \sum_j A_{ij}(e_i - e_j)$.

Definition 4.4. The weight of a TC-labeled graph G is:

$$w(G) = \prod_{\text{C-edge } (i \to j) \in G} D_{ij} \prod_{\text{T-edge } (i \to j) \in G} A_{ij}$$

Were ready to state the matrix-tree-cycle theorem.

Theorem 4.1. *The determinant* $\det(D + L)$ *is the signed weight-sum of tree-cyclic graphs:*

$$\det(D + L) = \sum_{\textit{tree-cyclic } G} (-1)^G w(G)$$

Proof. The i-th column of the matrix is $(D+L)_i = \sum_j D_{ij}e_j + \sum_j A_{ij}(e_i - e_j)$, so

$$\det(D+L) = \det\left(\begin{array}{c|c|c|c} D_{11}e_1 & D_{21}e_1 & & D_{n1}e_1 \\ + & + & & + \\ \vdots & \vdots & & \vdots \\ + & + & & + \\ D_{1n}e_n & D_{2n}e_n & & D_{nn}e_n \\ + & + & \cdots & + \\ A_{11}(e_1-e_1) & A_{21}(e_2-e_1) & & A_{n1}(e_n-e_1) \\ + & + & & + \\ \vdots & \vdots & & \vdots \\ + & + & & + \\ A_{1n}(e_1-e_n) & A_{2n}(e_2-e_n) & & A_{nn}(e_n-e_n) \end{array}\right) = \sum_{\text{TC-1-graph } G} w(G)\det(M_G) = \sum_{\text{tree-cyclic } G} (-1)^G w(G)$$

In the middle step we have again expanded the determinant by multilinearity: we sum over all possible ways to pick one term of each column. To choose a TC-labeled 1-graph on n vertices, is to choose for each vertex i (column i) an outgoing edge $i \to j$ and a label C (term $D_{ij}e_j$) or T (term $A_{ij}(e_i - e_j)$) for this edge. The final step is applying lemma 4.1. □

The matrix-tree theorem is recovered by taking D to be diagonal. In that case, the number of cycles and the number of C-edges is equal, so $(-1)^{\#\text{cycles}+\#\text{C-edges}} = 1$ and there are no signs involved. Each C-labeled self-loop corresponds to a root.

If we take $A = 0$ we get the fact that $\det(D)$ is the sum of signed cycle covers. A cycle cover of n vertices is a choice of edges such that each vertex has precisely one incoming and one outgoing edge, and its sign is the sign of the permutation that this graph depicts.

By taking $D_{ij} = 1$ for some i, j and zero elsewhere, we also ensure that the sign $(-1)^G = 1$, because in this case we have one C-edge and one cycle. Thus, $\det(D+L)$ will be the number of spanning trees rooted at i. The choice of j does not matter. By developing the determinant and using $\det(L) = 0$ we see that this is equal to $\det(L')$ where L' is obtained from L by deleting row i and column j.

By taking more off-diagonal entries of D to be nonzero, and by taking the corresponding rows and columns to be zero in A (and thus in L), we obtain the all-minors matrix-tree theorem.

Acknowledgements. Thanks to Darij Grinberg for pointing out mistakes and suggesting improvements, and for informing me about the relationship of Corollary 3.2.2 to Theorems 1&2 of [Rot01].

References

[Rot01] Gunter Rote. *Division-Free Algorithms for the Determinant and the Pfaffian: Algebraic and Combinatorial Approaches*, pages 119–135. Springer Berlin Heidelberg, 2001.