Symbolic Stochastic Systems and the Matrix Tree Theorem

Kumar Jeev and Jean Louis Lassez

Department of Computer Science
Coastal Carolina University
Conway, SC 29528-6054 , USA
{kjeev, jlassez}@coastal.edu

Abstract. Using symbolic Gaussian elimination we show that the Ergodic theorem for Markov chains is in fact a variant of the Matrix Tree theorem. This approach also provides insight into complexity issues regarding both theorems.

1 Motivation

Tarski's theorem on quantifier elimination provides the foundation for a main direction of automated theorem proving systems. In that respect it is a most powerful meta theorem, as it states that every theorem in elementary algebra and geometry can be proven by a single algorithm: quantifier (variable) elimination. However the inherent complexity of the algorithm imposes severe limitations to its systematic practical use. It was remarked in [3] that Tarski's result could nevertheless be used fruitfully for another type of theorem proving, called Qualitative Theorem proving. Instead of focusing on the constructive aspects of elimination, which lead to results of unmanageable size, one focuses on qualitative aspects of the results. As a consequence strikingly simple proofs of major theorems on the geometry of polytopes were found using solely elimination. (Gaussian elimination and Fourier/Motzkin elimination are particular instances of Tarski's elimination reduced to the linear case).

Here we look at another possible use of symbolic computation to prove theorems "by hand". We selected the Ergodic theorem of Markov Chains, it is a fundamental theorem in probability theory [10], and its proof makes use of principal eigenvectors and some topology to establish convergence and uniqueness of solution. One moves from state to state according to transitions probabilities, and the theorem aims at finding the probabilities of being in the various states after a certain length of time. It was interesting to see if we could prove it using solely Gaussian elimination. Running Maple [9] to generate a few examples, led us to conjecture the existence of a closed form symbolic solution. The size of the solution is indeed unmanageable, but we could characterize it simply in terms of spanning trees of the underlying Markov graph. We prove existence and unicity of the solution using (symbolic) Gaussian elimination, which was our initial aim, but we also have a combinatorial insight on the probabilities. Namely the probability to be in a state s increases with the number and weights of the reverse spanning trees having s as a root.

We can now see the Ergodic Theorem as a method to give us, for all states, the ratio of the number of reverse spanning trees having this state as a root and the total number of reverse spanning trees. But this is a variant of a fundamental theorem of Combinatorics , the Matrix Tree theorem, due to Kirchhoff [7], which gives the number of (reverse) spanning trees having a given root as the value of a determinant.

Applying once the Ergodic theorem we compute all ratios, applying the Matrix Tree theorem we compute the number for one vertex. From these we can compute the number of spanning trees for all vertices with two Gaussian eliminations, rather than n (number of vertices) if we applied the Matrix Tree theorem for each vertex.

We therefore claim three contributions: the first is a further validation of the claim that Tarski/Gauss/Fourier elimination can provide strikingly simpler and insightful proofs of major theorems. The second is the discovery that two fundamental theorems, one in probability and one in combinatorics are in fact essentially the same. The third contribution is a complexity result regarding the computation of number of spanning trees and the computation of steady state probability in a Markov Chain.

In the next section we discuss our versions of the Ergodic theorem and the Matrix Tree theorem and their implications. The third section contains the key lemma which has the essence of the common proof to both theorems. In the last section we prove the redundancy of the positivity constraints and the uniqueness of solution.

Both the Ergodic theorem and the Matrix Tree theorem are classical and can be found in standard texts in probability and combinatorics. We refer the reader to the notes [6] and the books [4, 5, 12] for background information.

2 The Ergodic Theorem and the Matrix Tree Theorem Revisited

A *Markov chain* is an exploration of a state space with the motion from one state to another being controlled by a set of transition probabilities. The Ergodic theorem forms the foundation of the theory of Markov chains [8, 10, 13]. It analyzes the steady state behavior of a Markov chain and lays down the conditions for the existence of a steady state solution to the probabilities of individual states. There are several versions of the Ergodic theorem and each one uses its own set of conditions under which a unique stationary solution exists. One such version of the Ergodic theorem states that “Any irreducible, finite, aperiodic Markov chain has all states Ergodic (reachable at any time in future) and has a unique stationary distribution, which is a probability vector.” [6, 12]

The transition probabilities of a Markov chain form the *transition matrix*, symbolized by M = $[p_{ij}]$, where p_{ij} is the probability of transition from state i to state j.

Definition 1. *A stochastic system is a system of constraints:* $M^T X = X$, $\sum x_i = 1$ *and* $x_i \geq 0$*, where M is a transition matrix.*

Construct a weighted digraph associated to the stochastic system, the *stochastic digraph*. The vertices represent the states and correspond to the variables, and there is an edge between each pair of vertices as well as a self-loop for each vertex. The transition probabilities are the weights of the edges of this digraph. The stochastic digraph is a complete digraph.

The *graph of a Markov chain* is a sub-graph of the stochastic digraph. It is the stochastic digraph without the zero weight edges.

An equation of the system is:

$$p_{1i}x_1 + p_{2i}x_2 + p_{3i}x_3 + \ldots + p_{ni}x_n = x_i = x_i(p_{i1} + p_{i2} + p_{i3} + \ldots + p_{in})$$

Informally, this left hand side of the equation represents "the flow" going out of a state while the right hand side represents "the flow" entering the state. Thus the system expresses a law of conservation for the long term state probabilities of a Markov chain. Observed for a sufficiently long time the probability of "getting into" any state is equal to the probability of "getting out" of it. The same conservation concept is used in Kirchhoff's laws of electricity and in the theory of networks. This leads us to a slightly generalized system of constraints.

Definition 2. *A conservation system for a matrix C=*$[c_{ij}]$ *where* $\forall i \forall j, c_{ij} \geq 0$ *and a diagonal matrix D=*$[d_{ij}]$*, is the system of constraints:* $C^T X = DX$, $d_{ii} = \sum_j c_{ij}$, $\sum x_i = 1$ *and* $x_i \geq 0$.

The graph of a conservation system without the zero weight edges is referred to as the *conservation digraph*. A stochastic system is a conservation system with D = I (Identity matrix).

Kirchhoff's Matrix Tree theorem is a fundamental tool in combinatorial theory [7]. It gives a method for evaluating the number of spanning trees in a graph. According to this theorem, "For an n-vertex digraph, define an n×n matrix A such that A[i,j] = 1 if there is an edge from i to j, for all i ≠ j, and the diagonal entries are such that the row sums are 0. Let A(k) be the matrix obtained from A by deleting row k and column k. Then the absolute value of the determinant of A(k) is the number of spanning trees rooted at k (edges directed towards vertex k)" [1, 2, 4, 5, 11] Thus A-D is the matrix which appears in the matrix tree theorem, where A is the adjacency matrix and D is the diagonal degree matrix.

A corresponds to the matrix C and D to the diagonal matrix of a conservation system. Thus the system of constraints: $A^T X = DX$, $x_i \geq 0$, and $\sum x_i = 1$ form a conservation system.

We will now state our main results, the proofs will be given in the following sections.

Definition 3. *A reverse spanning tree is a spanning tree whose edges' direction has been reversed.*

A reverse spanning tree of root r will be represented by the symbol T_r.

Definition 4. *The weight of a reverse spanning tree is obtained by multiplying all the transition probabilities associated to the edges of the reverse spanning tree.*

W(r) represents the sum of the weights of all reverse spanning trees with root r.

2.1 Ergodic Theorem revisited

If there exists a reverse spanning tree in the graph of the Markov chain associated to a stochastic system, then:

(a) the stochastic system admits the following probability vector as a solution:

$$\{x_i = \frac{W(i)}{\sum_j W(j)}\}_{i=1,n}$$

(b) the solution is unique.
(c) the conditions $\{x_i \geq 0\}_{i=1,n}$ *are redundant and the solution can be computed by Gaussian elimination.*

This is clearly a formulation of the Ergodic theorem for Markov chains which differs from the standard ones in the following ways:

The requirement that the graph of the Markov chain contains a reverse spanning tree covers various cases such as strongly connected, cycles, irreducible, and aperiodic found in the other formulations. It is also in a sense relevant to the goal of having solutions independent of the starting point of a random walk. Indeed it is easy to see that if the graph does not contain a reverse spanning tree, there will be several disconnected strongly-connected subgraphs and the long-term behavior of the random walk will depend on its starting point.

Another difference is interesting from a complexity point of view. The realization that the positivity conditions are redundant allows us to transform a linear programming problem into one of Gaussian elimination.

But the most striking difference comes from the fact that the probabilities are linked to numbers of reverse spanning trees in the Markov graph. So the Ergodic theorem can be reworded as a variant of the Matrix Tree theorem.

2.2 Matrix Tree Theorem revisited

If there exists a reverse spanning tree in a graph, then the ratios $\{x_i\}_{i=1,n}$ *of the number of reverse spanning trees with node i as their root to the total number of reverse spanning trees can be computed by solving the system* $A^T X = DX$, $\sum x_i = 1$, *where A is the adjacency matrix and D the diagonal degree matrix.*

It is a variant because the Matrix Tree theorem requires the computation of one determinant per root, while the variant computes all ratios of counts of spanning trees with one Gaussian elimination. So, in combination with the matrix tree theorem we can compute the number of spanning trees for all roots in just two Gaussian eliminations.

3 Symbolic Solution of a Conservation System

Definition 5. *The symbolic valuation of a reverse spanning tree is the symbolic product of the weights of the edges that belong to the tree.*

The sum of the symbolic valuations of all possible spanning trees with root r is represented by SV(r).

Lemma 1. *A symbolic solution of $C^T X = DX$, $d_{ii} = \sum_j c_{ij}$, and $\sum x_i = 1$ is*

$$\{x_i = \frac{SV(i)}{\sum_j SV(j)}\}_{i=1,n}$$

Proof:
The linear system is:

$$\begin{aligned}
c_{11}x_1 + c_{21}x_2 + c_{31}x_3 + \ldots + c_{n1}x_n &= x_1(c_{11} + c_{12} + c_{13} + \ldots + c_{1n}) \\
c_{12}x_1 + c_{22}x_2 + c_{32}x_3 + \ldots + c_{n2}x_n &= x_2(c_{21} + c_{22} + c_{23} + \ldots + c_{2n}) \\
c_{13}x_1 + c_{23}x_2 + c_{33}x_3 + \ldots + c_{n3}x_n &= x_3(c_{31} + c_{32} + c_{33} + \ldots + c_{3n}) \\
&\vdots
\end{aligned}$$

$$c_{1k}x_1 + c_{2k}x_2 + c_{3k}x_3 + \ldots + c_{nk}x_n = x_k(c_{k1} + c_{k2} + c_{k3} + \ldots + c_{kn}) \quad (1)$$

$$\begin{aligned}
&\vdots \\
c_{1n}x_1 + c_{2n}x_2 + c_{3n}x_3 + \ldots + c_{nn}x_n &= x_n(c_{n1} + c_{n2} + c_{n3} + \ldots + c_{nn})
\end{aligned}$$

an the normalizing equation is:

$$x_1 + x_2 + x_3 + \ldots + x_n = 1 \quad (2)$$

Let us replace the x_i by their proposed symbolic value in the system of equations:

$$\{x_i = \frac{SV(i)}{\sum_j SV(j)}\}_{i=1,n}$$

Equation (2) is trivially verified and equation (1) yields:

$$\sum_{r} c_{rk} SV(r) = \sum_{l} c_{kl} SV(k) \tag{3}$$

$\sum_r c_{rk} SV(r)$ on the left hand side of (3) corresponds to all the reverse spanning tree with r's as their roots and with an additional edge rk (directed from r to k) i.e. $\{T_r \bigcup rk\}_{r=1,n}$. Let us take one of these digraphs say G. The other state connected to k by an edge and leading towards the root r in G is denoted by l. G can also be thought of as representing a tree with k as the root and with an additional edge kl which is an element of $\{T_k \bigcup kl\}_{l=1,n}$. By construction the transformation is clearly a bijection between the set $\{T_r \bigcup rk\}_{r=1,n}$ and $\{T_k \bigcup kl\}_{i=1,l}$. Hence the proposed solution satisfies the system of equations.

If we set D=I making C a transition matrix, we obtain a *symbolic stochastic system* and we have as an immediate corollary a non-symbolic version of the lemma.

Corollary 1. *The probability vector* $\{x_i = \frac{W(i)}{\sum_j W(j)}\}_{i=1,n}$ *is a solution of the system* $M^T X = X$ *and* $\sum x_i = 1$ *if there exists a reverse spanning tree in the graph of the Markov chain.*

4 Uniqueness and Complexity

The proofs of the Ergodic theorem usually start from a previous result, the Chapman Kolmogorov equations, which compute a probability vector at time t+1 as the product by M^T of the probability vector at time t. Repeating this operation leads us, in general, to an eigen vector of the matrix M^T, satisfying the equation $M^T X = X$. But two problems arise: proving convergence as some cases, like circular Markov chains, do not converge and have to be treated as separate cases. And the other problem is uniqueness of the solution, which requires further mathematical machinery or restrictive hypotheses. In our approach these problems are removed, because of the requirement that the graph contains a reverse spanning tree. A slight generalization gives us the result for the variant of the Matrix Tree theorem.

Proposition 1. *The positivity condition in a conservation system is redundant if there exists a reverse spanning tree in the conservation digraph.*

Proof:
Let us assume without loss of generality that x_1 is negative. We know that,
$c_{21}x_2 + c_{31}x_3 + c_{41}x_4 + \ldots = (c_{12} + c_{13} + c_{14} + \ldots)x_1$

Since the right hand side is negative some of the variables on the left hand side also have to be negative. Let us assume that $x_2, x_3, \ldots, x_k$ are negative and

$x_{k+1}, \ldots, x_n$ are non-negative (because $\sum x_i = 1$ and all the variables cannot have negative values). Thus we have,

$c_{21}x_2 + c_{31}x_3 + ...c_{k1}x_k \leq (c_{12} + c_{13} + ...c_{1n})x_1$

Similarly from the rest of the equations of the conservation system with negative variables on the right hand side we get,

$c_{12}x_1 + c_{32}x_3 + \ldots + c_{k2}x_k \leq (c_{21} + c_{23} + \ldots + c_{2n})x_2$
$\ldots$
$c_{1k}x_1 + c_{2k}x_2 + \ldots + c_{(k-1)k}x_{k-1} \leq (c_{k1} + c_{k2} + \ldots + c_{k(k-1)} + c_{k(k+1)} + \ldots + c_{kn})x_k$

From which we derive,

$0 \leq [c_{1(k+1)} + c_{1(k+2)} + \ldots + c_{1n}]x_1 + [c_{2(k+1)} + c_{2(k+2)} + \ldots + c_{2n}]x_2 + \ldots + [c_{k(k+1)} + c_{k(k+2)} + \ldots + c_{kn}]x_k$

Similarly adding up the inequalities from the equations with the positive variables on the right hand side we get,

$0 \geq [c_{(k+1)1} + c_{(k+1)2} + \ldots + c_{(k+1)k}]x_{(k+1)} + [c_{(k+2)1} + c_{(k+2)2} + \ldots + c_{(k+2)k}]x_{(k+2)} + \ldots + [c_{n1} + c_{n2} + \ldots + c_{nk}]x_n$

The two inequations above are satisfied if and only if all the transition probabilities mentioned on the right hand side are zero. This means that there cannot exist an edge between a state with negative(positive) probability and a state with positive(negative) probability. This is impossible because there exists a reverse spanning tree in the conservation digraph. The assumption that there are negative variables in the solution to a conservation system does not hold.

Corollary 2. *A conservation system admits a unique solution, if there exists a reverse spanning tree in the graph of the associated Markov chain.*

Proof:
The set of solutions to a conservation system is bounded. Since there exists a reverse spanning tree we know that the system is solvable and that the conditions $\{x_i \geq 0\}_{i=1,n}$ are redundant. So, we have a non-empty bounded intersection of hyperplanes. This can happen only if they intersect in exactly one point.

As a consequence the solution to a conservation system can be computed by Gaussian elimination.

References

1. Biggs N., Algebraic Graph Theory, Cambridge University Press, Cambridge UK, 2nd edition, 1994.
2. Bondy J. A., Murty U. S. R., Graph Theory with Applications, North-Holland, New York, 1976.
3. Chandru V., Lassez J.-L., Qualitative Theorem Proving, International Symposium on Verification : Theory and Practice, Taormina, July 2003, Springer Verlag Lecture Notes in Computer Science.
4. Even S., Graph Algorithms, Computer Science Press, Maryland, 1979.
5. Golumbic M., Algorithmic Graph Theory and Perfect Graphs, Academic Press, 1980.
6. Karger D., Lecture Notes, 6.856J/18.416J Randomized Algorithms, MIT, 2002.
7. Kirchhoff G., Über die Auflösung der Gleichungen, auf welche man bei der untersuchung der linearen Verteilung galvanischer Ströme geführt wird, Ann. Phys. Chem. 72, 497-508, 1847.
8. Lukashin A. V., Borodovsky M., GeneMark HMM: New Solutions to Gene Finding, Nucleic Acid Research, 1998 Feb 15 26(4) 1107-1115.
9. Maplesoft, Maple 9 Student Edition, Waterloo Maple Inc., 2003.
10. Meyn, S. P. and Tweedie, R. L., Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993.
11. Moon J. W., Counting Labelled Trees, Canadian Mathematical Monographs, No.1, Canadian Mathematical Congress, 1970.
12. Motwani R., Raghavan P., Randomized Algorithms, Cambridge University Press, 1995.
13. Rabiner L. R., A tutorial on Hidden Markov Models, Proceedings of the IEEE, vol. 77, pp. 257-286, 1989.