First consider the case where $g_{i_1,\ldots,i_n}$ contains a spanning out tree t' rooted at k. Let σ be the permutation which maps $1,\ldots,n$ into that order in which these nodes occur in a breadth first traversal of t'. If i is encountered before j, there can be no edge from j into i unless i is the root since all indegrees are one. Therefore, since the edge into k is ignored by the corresponding matrix in the left side of (6) – the $k\,th$ row and column are there deleted – permuting rows and columns of this matrix according to σ yields an upper triangular matrix with diagonal elements $b_{v,u}$ for each edge $v \to u$ of t'.

The remaining case is that $g_{i_1,\ldots,i_n}$ contains no spanning out-tree rooted at k. Since each node has exactly one entering edge, tracing edges backwards will either connect an edge to the root – which if *every* node connects to k is contrary to hypothesis – or produces a cycle not including k. Therefore, let $\mathcal{C} = \{j_1,\ldots,j_\ell\}$ be the set of nodes involved in such a cycle. Note that the columns indexed by $\mathcal{C}$ in the matrix occurring in the left side of (6) corresponding to $g_{i_1,\ldots,i_n}$ are linearly dependent.

5 References

[1] J. B. Conway, *Functions of One Complex Variable*
Springer-Verlag, New York, 1978.

[2] A. Gibbons (1987), *Algorithmic Graph Theory.*
Cambridge University Press.

[3] H. Minc, *Nonnegative Matrices*
Wiley-Interscience, 1988.

[4] M.D.Vose, *Models of Genetic Algorithms*
University of Tennessee Technical Report CS-92-148, 1992

[5] Wilkinson, *The Algebraic Eigenvalue Problem*
Clarendon Press, Oxford 1965.

where all elements not shown are zero. Using the multilinearity of the determinant to expand the columns of (4) according to this decomposition yields (up to a constant)

$$\sum_{i_1 \neq 1} \cdots \sum_{i_n \neq n} \left| \begin{pmatrix} b_{i_1,1} & \vdots \\ \vdots & -b_{i_n,n} \\ -b_{i_1,1} & \vdots \\ \vdots & b_{i_n,n} \end{pmatrix}_{1k} \right| \tag{5}$$

Since the column sums of these matrices are zero, the kth *row* is the negative of the sum of the other rows. Making this substitution and expanding (by multilinearity of *rows*) yields (up to sign)

$$\left| \begin{pmatrix} b_{i_1,1} & \vdots \\ \vdots & -b_{i_n,n} \\ -b_{i_1,1} & \vdots \\ \vdots & b_{i_n,n} \end{pmatrix}_{1k} \right| = \left| \begin{pmatrix} b_{i_1,1} & \vdots \\ \vdots & -b_{i_n,n} \\ -b_{i_1,1} & \vdots \\ \vdots & b_{i_n,n} \end{pmatrix}_{kk} \right|$$

Using this relation in (5), canceling the exponential and logarithms in the left hand side of (3), and recalling that $B = Q^T$, we see that it suffices to show

$$\sum_{i_1 \neq 1} \cdots \sum_{i_n \neq n} \left| \begin{pmatrix} b_{i_1,1} & \vdots \\ \vdots & -b_{i_n,n} \\ -b_{i_1,1} & \vdots \\ \vdots & b_{i_n,n} \end{pmatrix}_{kk} \right| = \alpha \sum_{t \in \mathrm{Tree}_k} \prod_{u \to v \in \mathcal{E}(t)} b_{v,u} \tag{6}$$

where $\mathcal{E}(t)$ is the set of edges of t.

Next we show how the matrices – before removing the kth row and column – occurring in the left side of (6) correspond to graphs. The proof from this point parallels the argument given in [2]. If we interpret the occurrence of an off diagonal entry in row i and column j as a directed edge from i into j, then each matrix corresponds to a graph $g_{i_1,\ldots,i_n}$ where all indegrees (i.e., number of edges entering a node) are exactly one. This same interpretation applied to the matrix occurring in (4) – again, before removing row and column – yields a complete directed graph G over the nodes $\{1,\ldots,n\}$.

Note that each subgraph of G having all indegrees equaling one is represented by exactly one of the $g_{i_1,\ldots,i_n}$. In particular, each spanning *out-tree* of G rooted at k (i.e., all edges point *away* from the root) is represented by exactly $n-1$ of the $g_{i_1,\ldots,i_n}$; once for each of the possible $n-1$ edges pointing into k.

We next show that the determinant of the matrix corresponding to $g_{i_1,\ldots,i_n}$ occurring in the left side of (6) is nonzero if and only if $g_{i_1,\ldots,i_n}$ contains a spanning out-tree t' rooted at k, and in such case has value

$$\prod_{v \to u \in \mathcal{E}(t')} b_{v,u}$$

This will complete the proof since $v \to u$ is an edge of the out-tree t' if and only if $u \to v$ is an edge of the in-tree t which is obtained from t' by reversing edge directions.

The general framework of this example is a Markov chain whose $i \to j$ transition probabilities are converging to zero exponentially fast for $i \neq j$ (i.e., $\mathcal{A}$) and where the rates of exponential decay are essentially constant (i.e., $\mathcal{Q}$). The question of interest is what the steady state distribution is converging to. As we have shown, the answer is determined by the set S of minimal tributaries (corresponding to $\mathcal{Q}$).

This problem arises naturally when considering Markov chains converging to a dynamical system. A particular case of this general type is illustrated in [4] where the dynamics of Genetic Algorithms are analyzed.

4 Proof

Before proceeding, note that the theorem does provide the representation sought. Since entries of $\mathcal{Q}$ are probabilities, the edge weights $-\ln \mathcal{Q}_{i,j}$ are nonnegative and there can be no cancellation in computing the cost of a tributary. Because exponentials are nonnegative, there is likewise no cancellation in summing over tributaries.

In the case where $\mathcal{Q}$ has positive entries, the solution π to the steady state equation (1) is unique and has nonnegative entries (up to scaling) [3], and (2) is a nontrivial solution to (1) [5]. It therefore suffices in this case to show

$$\det\left((\mathcal{Q} - I)_{k1}\right) \;=\; \pm\,\alpha \sum_{t\,\in\,\mathrm{Tree}_k} \exp\{-\,r\,|\,t\,|\} \tag{3}$$

for some nonzero constant α which is independent of k. The general case follows by noting that when π is computed according to the theorem, the steady state equation represents an equality between polynomials in the positive variables $\mathcal{Q}_{i,j}$ and is therefore an algebraic identity [1].

Since determinants are insensitive to transpose, the left hand side of (3) is (up to sign)

$$\left| \begin{pmatrix} \sum\limits_{i\neq 1} b_{i,1} & -b_{1,2} & \dots & -b_{1,n} \\ -b_{2,1} & \sum\limits_{i\neq 2} b_{i,1} & \dots & -b_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -b_{n,1} & -b_{n,2} & \dots & \sum\limits_{i\neq n} b_{i,n} \end{pmatrix}_{1k} \right| \tag{4}$$

where $B = \mathcal{Q}^T$. Note that each *column* is a sum of vectors of the form

$$\begin{pmatrix} \vdots \\ \pm b_{i,j} \\ \vdots \\ \mp b_{i,j} \\ \vdots \end{pmatrix}$$

The results of this paper are to provide a solution to (1) in which no cancellation occurs, and to apply the resulting form to determine the behavior of the steady state distribution of a Markov chain as its transition probabilities decay exponentially. This situation arises naturally when considering Markov chains converging to a dynamical system.

2 Graphical Representation

A (finite) Markov chain can be represented by a complete directed weighted graph over $\{1, \ldots, n\}$ such that the $i \to j$ edge is labeled by a weight which encodes the i, jth entry of the transition matrix $\mathcal{Q}$.

We define a *tributary* of a complete directed weighted graph as a spanning *in-tree*, i.e., a tree containing every vertex of the graph and such that all edges point towards the root.

Let Tree_k be the set of tributaries rooted at k, and for $t \in \mathrm{Tree}_k$ let its cost $|t|$ be the sum of its edge weights. There is a beautiful connection between these tributaries and the steady state distribution of a Markov chain.

Theorem: Let $\mathcal{Q}$ be the $n \times n$ transition matrix for a Markov chain and let the corresponding graphical representation have edge $i \to j$ weighted by $-\ln \mathcal{Q}_{i,j}$. A solution to the steady state equation $x = x\mathcal{Q}$ has components given by

$$x_k \quad = \sum_{t \in \mathrm{Tree}_k} \exp\{-\,|t|\}$$

Note that if $\mathcal{Q}_{i,j}$ is zero, then the $i \to j$ edge is labeled by infinity, and any spanning tree containing it contributes nothing to the sum. Hence one need only consider the subgraph which has edges corresponding to nonzero entries of $\mathcal{Q}$. Moreover, self loops may also be ignored because trees contain no cycles.

3 Application

If we leave the graph (hence all weights and tributaries) referred to in the theorem unchanged, but modify the $i \to j$ transition probability to $\mathcal{A}_{i,j} = (\mathcal{Q}_{i,j})^r$, we obtain a Markov chain parametrized by r and with transition matrix $\mathcal{A}$ which has steady state solution

$$< \sum_{t \in \mathrm{Tree}_1} \exp\{-\,r\,|t|\} \;, \ldots, \sum_{t \in \mathrm{Tree}_n} \exp\{-\,r\,|t|\} >$$

where the Tree_k are computed from the graph representing $\mathcal{Q}$. Note that the minimum cost tributary – assuming it is unique – corresponds to the dominant term, and as $r \to \infty$ the steady state distribution corresponding to $\mathcal{A}$ therefore converges to point mass at that state at which it is rooted.

In the case where the set S of minimal tributaries is not a singleton, the limit of the steady state distribution of $\mathcal{A}$ is obtained (up to normalization) by setting r to 1 and restricting summation to range over S.

1 Introduction

This technical report concerns an independently discovered result, which (unfortunately for the author) was later discovered to be previously published:

> V.Anantharam & P. Tsoucas (1989), "A proof of the Markov chain tree theorem" *Statistics and Probability Letters*, vol. 8, pp. 189-192.

The technical report is made available because the proof of the Markov chain tree theorem is different, and because an application of the theorem to the type of situation which occurs when investigating Genetic Algorithms is outlined.

Markov chains are powerful tools for understanding stochastic algorithms. A well known example is given by the convergence theory for simulated annealing. A recent example where they also play a prominent role is given in [4] where the dynamics of Genetic Algorithms are analyzed.

The context with which we are interested here is given by regarding a Markov chain as a matrix which transforms probability distributions. Let $\mathcal{Q}$ be a nonnegative $n \times n$ matrix with row sums equaling 1, and let ψ^0 be a nonnegative row vector of dimension n which also sums to 1. The vector ψ^0 may be regarded as a probability distribution over the set $\{1, \ldots, n\}$, and $\mathcal{Q}$ is a transformer of probability distributions via the inductive definition

$$\psi^{t+1} = \psi^t \mathcal{Q}$$

Under well known conditions [3] the steady state distribution π defined by

$$\pi = \lim_{t \to \infty} \psi^t$$

exists, and in such case satisfies the steady state equation

$$\pi = \pi \mathcal{Q} \tag{1}$$

Numerical calculation of π is often no difficulty since fast reliable eigenvector routines are widely available. However, when the objective is theoretical analysis, and the transition matrix $\mathcal{Q}$ contains symbolic entries, a noniterative solution to (1) may be necessary to proceed.

Let I be the $n \times n$ identity matrix. If $\mathcal{Q} - I$ has rank $n - 1$ (which is typical in many applications), a solution is provided by determinants. Given a square matrix M, define M_{ij} to be that submatrix obtained from M by removing the ith row and jth column. When M has rank $n - 1$, a nontrivial element in its kernel is given by

$$< \det(M_{11}), -\det(M_{21}), \ldots, (-1)^{1+n} \det(M_{n1}) > \tag{2}$$

Hence by choosing M to be $\mathcal{Q} - I$, the vector (2) becomes a solution of (1).

The alternations in sign produced by naive expansion of these determinants can make identification of the dominant terms confusing. In particular, some terms one might expect do not in fact occur; they cancel out.

Computing Steady State Distributions
Without Cancellation[1]

MICHAEL D. VOSE

C.S. Dept., 107 Ayres Hall
The University of Tennessee
Knoxville, TN 37996
(615) 974-5067 vose@cs.utk.edu

Abstract: The steady state distribution of a Markov chain is represented such that no cancellation occurs. The resulting form is applied to determine the behavior of the steady state distribution of a Markov chain as its transition probabilities decay exponentially. This situation arises naturally when considering Markov chains converging to a dynamical system.

[1]This research was supported by the National Science Foundation (IRI-8917545) and AFOSR (90-0135).