Abstract of “An Algorithm to Compute the Stochastically Stable Distribution of a Perturbed Markov Matrix” by John R. Wicks, Ph.D., Brown University, August 2008.

Recently, some researchers have attempted to exploit state-aggregation techniques to compute stable distributions of high-dimensional Markov matrices (Gambin and Pokarowski, 2001). While these researchers have devised an efficient, recursive algorithm, their results are only approximate. We improve upon past results by presenting a novel state aggregation technique, which we use to give the first (to our knowledge) scalable, exact algorithm for computing the stochastically stable distribution of a perturbed Markov matrix. Since it is not combinatorial in nature, our algorithm is computationally feasible even for high-dimensional models.

An Algorithm to Compute the Stochastically Stable Distribution of a Perturbed Markov Matrix

by

John R. Wicks

B. S., Mathematics/Computer Science and Economics, Carnegie Mellon University, 1983

M. S., Mathematics, Carnegie Mellon University, 1983

S. M., Mathematics, University of Chicago, 1983

Ph. D., Mathematics, University of Chicago, 1990

Sc. M., Computer Science, Brown University, 2007

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy in the
Department of Computer Science at Brown University

Providence, Rhode Island
August 2008

Acknowledgements

I would like to thank my advisor, Prof. Greenwald, for all her support, Prof.s Cetintemel and Serrano for their willingness to serve on my dissertation committee, Alex Kruckman for his help proof-reading and editing, and my wife and children without whose love and understanding I could never have completed my degree.

Introduction

Probabilistic models pervade almost all areas of computer science today (e.g., computer vision, graphics, intelligent agents, and natural language processing). One common modeling tool is that of a finite-state, stationary Markov chain, which is characterized by an initial probability distribution and a (Markov) transition matrix that satisfies the Markov property. The long-term behavior of such a chain can be summarized by another probability distribution, which is a particular example of a *stable* distribution, or "mixed" equilibium. Under certain conditions, a Markov matrix has a unique stable distribution, which may then be computed using standard linear algebra techniques. In general, however, a Markov matrix may have an infinite number of mixed equilibria, so that determining the long-term behavior of the chain requires more difficult analysis.

Economists and game theorists have also used such models to study, for example, market dynamics and learning in repeated games. It is quite common for such models to have multiple mixed equilbria. Since individuals do not always behave *rationally* (i.e., optimally), some researchers have introduced an additional parameter, ϵ, that captures the "mistakes" (i.e., sub-optimal choices) that individuals sometimes make, which has the added benefit of forcing the model to a unique long-run equilibrium. The resulting model is called a *perturbed* Markov chain, and the corresponding transition matrix is then a *perturbed* Markov matrix (PMM), with entries that are *functions* of ϵ. Of particular interest is the limit of the stable distributions of a PMM as $\epsilon \to 0$, the so-called *stochastically stable* distribution (SSD) of a PMM (Kandori et al., 1993; Young, 1993), which is known to exist and is unique.

A naive approach to computing the SSD of a PMM is to simply to fix ϵ at a very small value and to compute the corresponding stable distribution of the resulting unperturbed Markov matrix using traditional linear algebra techniques. Repeating this computation for a decreasing sequence of ϵs yields a sequence of approximations to the SSD. However,

without precise analytic bounds on the error of such approximations (as a function of ϵ), they do not really say anything about the SSD. An exact combinatorial algorithm for computing the SSD is known (Friedlin and Wentzell, 1984), but it involves enumerating certain spanning subtrees of the graph associated with the PMM. Because sufficiently expressive Markov models tend to be very high-dimensional, and because the number of spanning subtrees grows exponentially with the dimension, such an approach is not feasible in general.

Recently, Gambin and Pokarowski (2001) have attempted to exploit state-aggregation techniques to compute stable distributions of high-dimensional Markov matrices. While these researchers have devised an efficient, recursive algorithm, their results are only approximate. We improve upon past results by presenting a novel state aggregation technique, which we use to give the first (to our knowledge) scalable, exact algorithm for computing the stochastically stable distribution of a perturbed Markov matrix. Since it is not combinatorial in nature, our algorithm is computationally feasible even for high-dimensional models. Researchers in economics have already used our approach to study the dynamics of housing markets. Given the widespread use of Markov models in computer science, we imagine that it will soon find direct applications there, as well.

Overview

This thesis is divided into three parts. Part I focuses on Markov matrices and their stable distributions. This part sets the groundwork for Part II, on perturbed Markov matrices (PMMs) and their stochastically stable distributions (SSDs). It is here where we present our algorithm for computing the SSD of a PMM. Part III presents two additional algorithms, which were inspired by our algorithmic work on computing the SSD of a PMM; however, making these theoretical connections precise remains for future work.

Overview of Part I

In more detail, the main goal of Part I is to introduce our novel approach to state aggregation in a Markov chain, which we call *reduction*, given in chapter 5. Unlike related techniques, reduction actually "eliminates" states from consideration by compressing time. In fact, state aggregation is only a side-effect of reduction that arises when we choose to eliminate a maximal number of states.

In Part II, we show that reduction can be generalized to PMMs in a manner that is amenable to (real) analysis. While the primary goal of Part I is to introduce the construction of chapter 5 for use in Part II, we will illustrate its usefulness immediately (in Part I) by proving a number of "structure" theorems for Markov matrices. That is, we will use the construction to develop novel proofs of classic results on the nature of the set of stable distributions of a Markov matrix.

We first present our reduction construction in the context of Markov *matrices*, deferring making the connection with Markov chains until later. Although the construction may be defined algebraically, the intuition behind it is geometric. So we begin with a combination of graph theory and linear algebra in chapter 1, showing how algebraic properties of a

Markov matrix, M, may be expressed in terms of its associated graphs, where the vertices of the graph correspond to indices of M.

Two key concepts defined in chapter 1 are *open* and *closed* sets of vertices in a graph. We will show, for example, that there always exists a walk from any vertex in open set that exits that set, and that we can apply our construction to eliminate a set s of indices of M iff s corresponds to an open set of vertices in the graph of M. We will also begin to demonstrate the connection between closed sets of M and its collection of stable distributions.

Because the reduction construction is defined in terms of submatrices, in chapter 2, we develop sufficient theory to carefully define and analyze the behavior of certain submatrices of a Markov matrix. In chapter 3, we give a novel proof that $M^\infty \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} M^j$ exists for any Markov matrix, M. This will allow us to prove the first of our structure theorems, characterizing the set of stable distributions of a Markov matrix. Next, in chapter 4, we give an algebraic characterization of open sets, which prepares us for chapter 5.

Finally, in chapter 5, we present our main construction, *reduction*, along with another important one we call *scaling*. For unperturbed matrices, scaling may be recognized as right-preconditioning, a standard technique used to solve linear systems of equations. Reduction is more subtle, in that it will allow us to "eliminate" open sets of indices.

In this chapter we also define two novel notions of *equivalence* between Markov matrices, and show that we may recover the set of stable distributions of a given Markov matrix from the corresponding set of any equivalent one. This is a non-trivial result, in that, even though reduction produces a Markov matrix of strictly smaller dimension, we can still prove that the result is, in a precise sense, equivalent to the original. Thus, if we are only interested in computing stable distributions, reduction is a powerful tool for simplifying high-dimensional Markov models.

Overview of Part II

The heart of Part II is our algorithm for computing the SSD of a PMM, presented in chapter 7, section 7.7. However, we begin by taking time to prove the Markov Chain Tree Theorem (MCTT) in detail. As we will explain in the following paragraph, this theorem relates the stable distributions of a Markov matrix to the collection of directed spanning subtrees of its associated graph. Although it is not computationally practical, the MCTT

provides the theoretical basis for most of chapter 7. We present a novel proof of the MCTT, which exploits basic properties of the determinant and which, we feel, has a pleasing geometric flavor to it.

Thus, in chapter 6, we will establish some geometric preliminaries on directed spanning trees. We will show that there always exists a walk from any vertex that enters some closed set of vertices, and in particular, if a graph contains exactly one closed set, then it contains a directed, spanning subtree rooted at each vertex of that set. We will then define a vector, w_M, which, in the case that M is unichain (i.e., its associated graph has a single closed class), will turn out to be proportional to its unique stable distribution. We will prove that this is the case by defining another vector in terms of determinants (specifically, as the diagonal of the *adjoint* of the *laplacian*, $M - I$), which is easily seen to be proportional to that stable distribution, as well as to w_M.

In chapter 7, we move on to give a precise definition of a perturbed Markov matrix, M_ϵ, and its associated stochastically stable distribution. The key issue throughout the chapter is that we must be able to take limits as ϵ goes to 0 (i.e., continuity). Thus, the entries of M_ϵ must be sufficiently well-behaved, they must remain so as such as we operate on M_ϵ, and, for sufficiently small ϵ, the entries in M_ϵ are either identically zero or positive. These three conditions on the analytic nature of M_ϵ, effectively force the entries of a PMM to be in a certain class of functions, known as *exponentially convergent* functions.

Continuity is an obvious restriction. Although somewhat vaguely stated, at this point, since we will want to perform standard linear algebraic operations on M_ϵ, the second condition is also plausible. The third condition is a bit more subtle. As suggested in the Introduction, M_ϵ is supposed to have a unique stable distribution (i.e., mixed equilibrium) for small, positive ϵ. This corresponds to the property that M_ϵ be *unichain* for sufficiently small $\epsilon > 0$. This property is defined in terms of the (unweighted) graph (cf. section 1.2) associated with M_ϵ (for each fixed value of $\epsilon > 0$), i.e., that it has a unique closed class. For consistency, it is reasonble to require that this *unweighted* graph not change as we vary ϵ. This corresponds exactly to our third requirement on the entries of a PMM. This allows us to define the unweighted graph associated with M_ϵ, which we will denote by $G_-(M_\epsilon)$. and not worry about whether this refers to the graph of the PMM or the graph of the Markov matrix for a fixed ϵ, since these must be equal.

Thus, in section 7.1, we introduce the class of exponentially convergent functions and

discuss how members of this class behave both algebraically and analytically. After first defining *perturbed* matrices in section 7.2 (as matrices with exponentially convergent entries), we define perturbed *Markov* matrices PMMs in section 7.3. Using the MCTT, we show that the unique stable distribution of M_ϵ is a perturbed matrix, so that its limit as $\epsilon \to 0$, i.e., the stochastically stable distribution of M_ϵ, is well-defined. In sections 7.4-7.6, we show how the concepts of equivalence, scaling, and reduction from chapter 5 generalize to PMMs.

We then use these constructions, in section 7.7, to give our algorithm for computing the SSD of a PMM. The two fundamental difficulties with designing such an algorithm are:

- how to efficiently represent a PMM for algebraic computation, and
- how to carry out the necessary algebraic computations *without ever inverting a PMM*.

By a careful appeal to the MCTT, we show that:

- we may represent any PMM by a pair of real-valued matrices, and
- by applying reduction to eliminate open sets with respect to $M_0 \equiv \lim_{\epsilon \to 0} M_\epsilon$, we need only invert submatrices of M_0, i.e., unperturbed Markov matrices.

Finally, in order to guarantee that our algorithm makes progress and eventually terminates, we use scaling in a rather subtle manner (cf. Corollary 7.17).

Overview of Part III

Part III presents two additional algorithms, which were inspired by our algorithmic work on computing the SSD of a PMM.

In chapter 8, we reformulate the problem of topologically sorting a directed graph, usually restricted to directed, *acyclic* graphs, as a multi-objective optimization problem over *arbitrary*, weighted, directed graphs. We present an algorithm and prove that it yields an optimal weighted, topological sort. When combined with suitable empirical techniques for generating meaningful graphs, this algorithm could yield interesting results in several application domains, including ranking, preference aggregation, and information retrieval. As such, we have dubbed our algorithm **GraphRank**. We conjecture that a variation of

the SSD algorithm, tailored to only compute the "exponents" of the stable distribution, will compute precisely the same solution as **GraphRank**. If true, this would give us a Markov chain interpretation for our ranking solution, a la Dwork et al. (2001) In brief, we would model each edge of the original graph as an individual PMM, and represent the entire graph by a convex combination of all such PMMs. We conjecture that the "exponents" of the stable distribution of the resulting PMM yield the same optimal, weighted topological sort of the original graph as produced by **GraphRank**.

In chapter 9, we present another algorithm ranking algorithm, which we call **QuickRank**. This algorithm is recursive, and can be used to rank individuals in social network, based on an associated hierarchy. For example, these individuals may be research articles, in which case the social network would be given by citations, and the hierarchy specified by areas of specialization.

This is actually not simply a single algorithm, but a whole class of algorithms parameterized by a given "base" ranking algorithm, which we apply at each level in the hierarchy. One view of our approach is that it suitably modifies a given "base" ranking algorithm so that the resulting ranking satisfies two intuitively desireable axioms, which we have dubbed *peer review* and *Bonacich's hypothesis*.

The idea of exploiting a hierarchy in this way has been suggested in in previous work. For example, using only a 2-level hierarchy, determined essentially by URL domains, with PageRank as the base ranking algorithm, yields the BlockRank algorithm of Kamvar et al. (2003b). It should be noted that they did not argue to use the results of BlockRank directly, but only as a good, first approximation to PageRank. In contrast, we argue that the ranking scheme of **QuickRank** may produce superior results, in that they more acurately reflect the judgements of "local" experts and are resistant to the ranking manipulation technique of "web-spamming".

As with **GraphRank**, we conjecture that, when the base ranking algorithm has a Markov chain interpretation (such as in the case of PageRank), the ranking which **QuickRank** produces may characterized as the SSD of an appropriately framed PMM. While we have been able to do this for a 2-level hierarchy, this remains an open problem for general hierarchies.

Contents

List of Tables

List of Figures

Part I

Markov Matrices

Chapter 1

Markov Matrices and "Markov" Graphs

In this chapter, we compile a collection of definitions and facts regarding Markov matrices and their associated graphs. Throughout, we will rely on the following notation. We let S_n denote the set of integers from 1 to n, and we let S_n^0 denote the set of integers from 0 to n. Often, S_n will represent the index set for an $n \times n$ square matrix. We will also use S_n and S_n^0 to define sequences, where $\sigma : S_l^0 \to S_n$ defines a sequence on S_n of length $l+1$. We will denote the i^{th} element of σ by σ_i (instead of $\sigma(i)$).

1.1 Graph Theory Essentials

We begin with some basic notions from graph theory. Specifically, we will:

- give formal definitions of (un)directed, (un)weighted graphs,
- define (strongly) connected components of a graph, as well as, open and closed sets of vertices, and
- state and prove some basic properties of open and closed sets of vertices that we will need in subsequent chapters.

1.1.1 Basic Definitions

We will define a *directed graph*, $G = (V, E, s, t)$, as a 4-tuple in which V is a set of *vertices*, E is a set of *edges*, and $s : E \to V$ and $t : E \to V$ are mappings from edges

to vertices. We will restrict our attention to graphs in which both V and E are finite. In this figure, vertices are drawn as geometric points, and edges as arrows from one vertex to another. Specifically, the arrow corresponding to an edge, $\alpha \in E$, starts at the point $s(\alpha)$ and terminates (i.e., ends) at $t(\alpha)$. If $s(\alpha) = t(\alpha)$, then α is called a *self-loop*.

We will define an *undirected* graph as a directed graph, $G = (V, E, s, t)$, with the property that it contains the "reverse" of every edge. That is, for each $\alpha \in E$, there exists an $\alpha' \in E$ such that $s(\alpha) = t(\alpha')$ and $t(\alpha) = s(\alpha')$. Intuitively, we may view the pair α and α' as a single "composite" edge, drawn as an arrow with arrowheads on both ends, or alternatively, as a line segment with no arrowheads at all.

When there are no *repeated* edges in a graph G, (i.e., when there are no two edges, $\alpha_i \in E$, with the same starting and ending points, $s(\alpha_1) = s(\alpha_2)$ and $t(\alpha_1) = t(\alpha_2)$), we can represent E by the set of ordered pairs, $\{(s(\alpha), t(\alpha)) \in V \times V \mid \alpha \in E\}$. In this case, s and t are just the respective projections onto the first and second coordinates of each edge, and we can refer to the graph simply as $G = (V, E)$, with each edge represented as a pair of vertices. The order of this pair matters only when the graph is directed.

A *walk* of length l in a (directed or undirected) graph is a sequence of l edges, $\{\alpha_i\}_{i=1}^{l}$ such that $t(\alpha_i) = s(\alpha_{i+1})$ for $1 \leq i \leq l-1$. The walk starts at $s(\alpha_1)$ and ends at $t(\alpha_l)$. A *path* is a walk that does not revisit any edges or vertices, i.e., $\{\alpha_i\}_{i=1}^{l}$ is a path iff $l = |\{\alpha_i \mid 1 \leq i \leq l\}|$ and $l + 1 = |\{s(\alpha_i) \mid 1 \leq i \leq l\} \cup \{t(\alpha_i) \mid 1 \leq i \leq l\}|$. Note that since a path is a walk, and since we can always drop edges from a walk to obtain a path, there is a walk from v to w iff there is a path from v to w.

In a graph that has no repeated edges, a walk may also be specified by a sequence $\sigma : S_l^0 \to S_{|V|}$ of $l + 1$ vertices. Here, $v_{\sigma_0} = s(\alpha_1)$ is the first vertex in the walk and $v_{\sigma_l} = t(\alpha_l)$ is the last, with $(v_{\sigma_i}, v_{\sigma_{i+1}}) \in E$ and $v_{\sigma_i} = t(\alpha_i) = s(\alpha_{i+1})$, for all $1 \leq i \leq l-1$. In this definition, a path is a walk for which the corresponding σ is 1-1 (i.e., distinct inputs map to distinct outputs), so that no vertex is revisited. Note that this second definition is more general than the first (for graphs that have no repeated edges), since it allows walks (and paths) of length 0, which we specify by a single vertex, $\{v_i\}$. When we encounter such walks (and paths), we will say that they both start and end at v_i.

A *cycle* is a path of length 1 or more with the additional condition that the initial and final vertices are the same; that is, $\sigma_0 = \sigma_l$, or $s(\alpha_1) = t(\alpha_l)$. A *self-loop* is a cycle of length 1.

Given a directed graph, $G = (V, E)$, its *transitive closure*, $G_T = (V, E_T)$, is defined such that $(v_i, v_j) \in E_T$ iff there is a directed walk (or path) from v_i to v_j in G. Because we allow walks of length 0, $(v_i, v_i) \in E_T$ for all i. This definition allows us to define a natural preorder[1] on V, given by the "leads to" relation, $\rightsquigarrow$, where $v_i \rightsquigarrow v_j$ iff $\left(v_i, v_j\right) \in E_T$. This preorder gives rise to an equivalence relation, $\sim$, where $i \sim j$ (read, "v_i is *strongly connected* to v_j") iff $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Equivalence classes with respect to $\sim$ are often called the *strongly connected components* (SCCs) of G.[2] Note that SCCs are maximal, meaning they do not contain other SCCs; further, the SCCs of G *partition* the vertices of G, meaning each vertex belongs to exactly one SCC.

Similarly, we have a "connects to" relation, $\leftrightsquigarrow$, associated with the undirected graph corresponding to G. That is, $v_i \leftrightsquigarrow v_j$ iff there is an *undirected* walk (or path) from v_i to v_j in G. The equivalence classes associated with this relation are called the *connected components* of G. A *tree* is a connected component with no cycles. A graph is called *complete* if there is an edge from every vertex to every other vertex. A complete graph consists of only one connected component, which is in fact strongly connected.

Thus far, we have restricted our attention to directed and undirected *unweighted* graphs. Much of this thesis is actually concerned with weighted graphs. A weighted (directed or undirected) graph is one augmented with a function $d : E \to \mathbb{R}$, which assigns a real-valued "weight" to each edge in the graph. The weight of an edge, $d(\alpha)$, is drawn as a label on the corresponding arrow, and can be thought of as a cost or a likelihood of traversing α. Sometimes, we will be given a weighted graph $G = (V, E, d)$, but will wish to refer to the corresponding unweighted graph. To do so, we will use the notation, $G_- = (V, E)$.

Finally, we define the *restriction* of a (directed/undirected, weighted/unweighted) graph, G, with vertex set, V, and edge set, E, as follows. Given a set of vertices, $V' \subset V$, $G|_{V'}$ will be the subgraph with vertex set, V', the set E' of all edges with both ends in V', and the corresponding restrictions of all other ancillary functions (e.g., d).

[1] A *preorder* is a reflexive (i.e., $v \rightsquigarrow v$) and transitive (i.e., $u \rightsquigarrow v$ and $v \rightsquigarrow w$ implies $u \rightsquigarrow w$) relation.

[2] Strongly connected components may also reasonably be called *communicating classes*, to conform with the literature on Markov chains (see section 5.3).

1.1.2 Open Sets and Closed Classes

We will say that a subset of vertices, $V' \subset V$, is *invariant* iff V' has no outgoing edges, i.e., for all $\big(v_i, v_j\big) \in E$, if $v_i \in V'$, then $v_j \in V'$. An invariant SCC is referred to as a *closed* class. If V' does not contain a closed class, we will say that V' is *open*. Vertices that are do not belong to a closed class are called *transient*.

Note that the terms "open" and "closed" are not opposites here. "Closed" refers only to single (invariant) SCCs, while "open" can refer to a set of vertices larger than a single SCC. In fact, the vertices in an open set need not even be connected. However, any single SCC is either open or closed. If it is not open, it contains some closed class, which must be the entire SCC since SCCs are maximal, and so it is closed. If it is not closed, it cannot contain a closed class since SCCs are maximal, and so it is open.

We prove two simple lemmas in this section. The first is an intuitive observation about closed classes, namely that there is always a walk entering and terminating in a closed class. It follows immediately from this fact that every directed graph contains a closed class. The second is an intuitive observation about open classes, namely that there is always a walk exiting an open class. This second lemma follows as a simple consequence of the first.

Lemma 1.1. *Starting from any vertex in a directed graph G, there exists a walk terminating in a closed class. In particular, every directed graph contains a closed class.*

Proof. Let $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ be the SCCs of G. Pick an arbitrary vertex v, and call its SCC $\mathcal{C}_{\sigma_1}$. If $\mathcal{C}_{\sigma_1}$ is closed, then we have a walk (of length 0) starting at v and terminating in a closed class, and we are done. Otherwise, $\mathcal{C}_{\sigma_1}$ is open, and there is an outgoing edge (s_1, t_1) with $s_1 \in \mathcal{C}_{\sigma_1}$ and $t_1 \in \mathcal{C}_{\sigma_2}$ for some $\sigma_2 \neq \sigma_1$. Now since v and s_1 are in the same SCC, there is a walk from v to s_1, and continuing along the edge (s_1, t_1), there is a walk from v to t_1.

As above, if $\mathcal{C}_{\sigma_2}$ is closed, there is a walk terminating in a closed class, and we are done. Otherwise, we can repeat the process and find an outgoing edge (s_2, t_2) with $s_2 \in \mathcal{C}_{\sigma_2}$ and $t_2 \in \mathcal{C}_{\sigma_3}$ for some $\sigma_3 \neq \sigma_2$. Now since t_1 and s_2 are in the same SCC, there is a walk from t_1 to s_2, and continuing along the edge (s_2, t_2), there is a walk from v to t_2. Proceeding inductively, we either encounter a closed class, in which case we have found a walk from v terminating in a closed class and we are done, or we continue the sequence σ of open SCCs, and the walk from v to the vertices in each of these SCCs. In general, $s_i \in \mathcal{C}_{\sigma_i}$, $t_i \in \mathcal{C}_{\sigma_{i+1}}$, and $\mathcal{C}_{\sigma_i} \neq \mathcal{C}_{\sigma_{i+1}}$.

Suppose we never encounter a closed class. Since there are only finitely many SCCs in G, at some point σ must include a SCC that it had already included. So for some $i < j$, we have that $\sigma_i = \sigma_j$. There is a walk from v to s_{j-1}, the starting vertex of the incoming edge to $\mathcal{C}_{\sigma_j}$, that has two halves: a walk from v to s_i, the starting vertex of the outgoing edge from $\mathcal{C}_{\sigma_i}$, and a walk from s_i to s_{j-1}. Specifically, $s_i \rightsquigarrow s_{j-1}$. But $s_{j-1} \rightsquigarrow t_{j-1}$ (because of the edge (s_{j-1}, t_{j-1})), and $t_{j-1} \rightsquigarrow s_i$ (because $t_{j-1} \in \mathcal{C}_{\sigma_j} = \mathcal{C}_{\sigma_i}$), so $s_{j-1} \rightsquigarrow s_i$. In particular, $s_i \sim s_{j-1}$, which is a contradiction, since $\mathcal{C}_{\sigma_{j-1}} \neq \mathcal{C}_{\sigma_j} = \mathcal{C}_{\sigma_i}$.

As an immediate corollary, for any directed graph G, the fact that there is a walk from any vertex v that terminates in a closed class implies that G contains at least one closed class. □

Lemma 1.2. *A subset $V' \subset V$ of vertices in a directed graph, $G = (V, E)$, is open iff for every $v \in V'$ there is a walk from v to some vertex $z \notin V'$.*

Proof. Assume that V' is open, and consider an arbitrary vertex, $v \in V'$. By Lemma 1.1, there is a walk from v to some vertex, w, in some closed class, $\mathcal{C}$. Since V' is open, $\mathcal{C} \not\subset V'$. Choose $z \in \mathcal{C} \setminus V'$. Since w and z are in the same SCC, there is a walk from w to z. Hence, there is a walk from v to $z \notin V'$.

Now assume that V' is not open, i.e., that it contains a closed class, $\mathcal{C}$. We must produce a $v \in V'$ for which no walk in G from v leaves V'. We can choose any $v \in \mathcal{C}$. Since there is no edge leaving $\mathcal{C}$, there can be no walk from v that leaves $\mathcal{C}$, much less V'. □

1.1.3 Closed Classes in Subgraphs

In this section, we examine the relationship between the closed classes of a graph and the closed classes of certain subgraphs and restrictions. In particular, we observe that the number of closed classes in a graph cannot decrease as we remove its edges. In other words, the number of closed classes cannot increase as we add new edges. This observation will be particularly relevant in Chapter 7.

Lemma 1.3. *Given a directed graph $G = (V, E)$ with $V' \subset V$ an invariant set of vertices, if $\mathcal{C}'$ is a closed class of $G' = G|_{V'}$, then it is also a closed class of G.*

Proof. First, we will show that $\mathcal{C}'$ is invariant in G. By assumption, V' is invariant. So there are no edges in G starting at vertices inside $\mathcal{C}'$ and ending at vertices *outside* V'. It

remains to show that there are no edges in G starting at vertices inside $\mathcal{C}'$ and ending at vertices *inside* V'. Since $\mathcal{C}'$ is closed in G', there are no such edges in G'. Further, since G' is a restriction of G, their edge sets coincide on the restricted set of vertices. Hence, there cannot be any such edges in G either, and $\mathcal{C}'$ is invariant in G.

Second, we must show that $\mathcal{C}'$ is a SCC of G, that is, for all $v \in \mathcal{C}'$, $w \in V$, $v \sim w$ in G iff $w \in \mathcal{C}'$. If $v \sim w$, then w must be in $\mathcal{C}'$, because $\mathcal{C}'$ is invariant in G. Conversely, if $w \in \mathcal{C}'$, then $v \sim w$ in G', so v must also be strongly connected to w in G, since any edge in G' is also in G. Therefore, $\mathcal{C}'$ is a closed class of G. □

Lemma 1.4. *Any closed class $\overline{\mathcal{C}}$ in a connected component $\overline{G} = (\overline{V}, \overline{E})$ of a directed graph G is also a closed class of G.*

Proof. Connected components have no incoming or outgoing edges, so $\overline{V}$ is invariant, and the restriction $G|_{\overline{V}}$ is exactly $\overline{G}$. Hence, we can apply Lemma 1.3 with $G' = \overline{G}$ and $\mathcal{C}' = \overline{\mathcal{C}}$ to conclude that $\overline{\mathcal{C}}$ is a closed class of G. □

Lemma 1.5. *If $\overline{G} \subset G$ with $\overline{V} = V$ and $\overline{E} \subset E$, then every closed class of G contains some closed class of $\overline{G}$.*

Proof. Let $\mathcal{C}$ be a closed class of G, and consider $\overline{G}\big|_{\mathcal{C}}$. (Note that this restriction is not well-defined unless $\overline{V} = V$.) By Lemma 1.1, $\overline{G}\big|_{\mathcal{C}}$ contains a closed class, call it $\overline{\mathcal{C}}$. By construction, $\overline{\mathcal{C}}$ is contained in $\mathcal{C}$, so we have only to argue that $\overline{\mathcal{C}}$ is closed in $\overline{G}$. Because $\mathcal{C}$ is invariant in G, it is also invariant in $\overline{G}$, since $\overline{E} \subset E$. Hence, we can apply Lemma 1.3, with $G = \overline{G}$, $G' = \overline{G}\big|_{\mathcal{C}}$, and $\mathcal{C}' = \overline{\mathcal{C}}$, to conclude that $\overline{\mathcal{C}}$ is a closed class of $\overline{G}$ which is contained in $\mathcal{C}$. □

1.2 Markov Matrices

We will now introduce our fundamental objects of study, Markov matrices and their stable distributions. Specifically, we will:

- define a Markov matrix, its laplacian, and its set of stable distributions;
- associate a weighted directed graph with any principal submatrix of a Markov matrix (i.e., a sub-Markov matrix); and

- carry over the graph-theoretic concepts of Section 1.1 to Markov matrices in order to define irreducible and unichain Markov matrices.

An $m \times n$ *matrix* M has m rows and n columns. We write $M_{i,j}$ to refer to the element in the i^{th} row and j^{th} column of M. Observe that $M_{i,j} = e_i^t M e_j$, where $(e_i)_j = [i = j]$,[3] i.e., e_i has a 1 in the i^{th} component, and 0s elsewhere.

Two special cases of matrices arise when one of the *dimensions* (either m or n) equals 1. Specifically, a *column* vector is an $n \times 1$ matrix; likewise, a *row* vector is an $1 \times n$ matrix. To keep our notation brief, we will index vectors using one variable instead of two. That is, for a column vector v, $v_i = v_{i,1}$, and for a row vector w, $w_j = w_{1,j}$. The set of column vectors of dimension $n \times 1$ comprise the vector space $\mathbb{R}^n$. Unless otherwise specified, when we say "vector," we mean a column vector in $\mathbb{R}^n$.

A *submatrix* of an $n \times n$ matrix, M, is obtained by eliminating $m < n$ rows and columns of M to obtain an $(n - m) \times (n - m)$ matrix. A submatrix is called *principal* if the set of removed rows is the same as the set of removed columns.

We denote the l_1-norm on $\mathbb{R}^n$ by $\|\cdot\|_1$. For any $v \in \mathbb{R}^n$, this is the sum of the absolute values of its entries:

$$\|v\|_1 = \sum_{i=1}^{n} |v_i| .$$

We will use the same notation to denote the corresponding induced matrix norm on the set of $n \times n$ matrices with real-valued entries,

$$\|M\|_1 = \max \{\|Mv\|_1 \mid v \in \mathbb{R}^n, \|v\|_1 = 1\} ,$$

and we will take as known the fact (Horn and Johnson, 1985, p. 294) that

$$\|M\|_1 = \max \left\{ \sum_{i=1}^{n} \left|M_{i,j}\right| \mid 1 \leq j \leq n \right\} .$$

In other words, $\|M\|_1$ is the maximum of the column sums of the absolute values of the entries of M. As a matrix norm, $\|\cdot\|_1$ obeys the triangle inequality (i.e., $\|M + N\|_1 \leq \|M\|_1 + \|N\|_1$) and is sub-multiplicative (i.e., $\|MN\|_1 \leq \|M\|_1 \|N\|_1$).

We denote the set of all $n \times n$ square matrices with non-negative, real-valued entries by $\mathrm{Mat}_n\left(\mathbb{R}^+\right)$. A matrix $M \in \mathrm{Mat}_n\left(\mathbb{R}^+\right)$ is called *Markov* iff $JM = J$, where $J =$

[3] We use Iverson's convention: for any proposition, Q, $[Q] = 1$, if Q is true, and 0 otherwise (Knuth, 1997, p. 32).

$(1, \ldots, 1)$ ambiguously denotes a row vector of 1s of arbitrary length. In other words, all columns in a Markov matrix sum to 1. Observe that: for any Markov matrix, M, $\|M\|_1 = 1$; likewise, for any submatrix M' of M, $\|M'\|_1 \leq 1$.

We will sometimes refer to a principal submatrix, $\overline{M}$, of a Markov matrix, M, as a *sub-Markov* matrix. Given a sub-Markov matrix, $\overline{M}$, we can define its *laplacian*, $\Lambda\left(\overline{M}\right) \equiv \overline{M} - I$. By convention, we will abbreviate $\Lambda\left(\overline{M}\right)$ as $\overline{\Lambda}$, $\Lambda\left(M_1\right)$ as Λ_1, etc.

Notice that if M is Markov, then:

- $\Lambda_{i,j} = M_{i,j} \geq 0$, if $i \neq j$, i.e., Λ has non-negative off-diagonal entries;
- for any j, $\sum_{i \neq j} \Lambda_{i,j} = \sum_{i \neq j} M_{i,j} = 1 - M_{j,j} \leq 1$, i.e., Λ's off-diagonal column sums are less than or equal to 1; and
- $J\Lambda = JM - JI = J - J = 0$, i.e., Λ's columns sum to 0.

Conversely, it is easy to check that if Λ satisfies these three conditions, then $M = \Lambda + I$ is Markov.

For any matrix M and vector v, if $Mv = \lambda v$, we say that v is an *eigenvector* of M with *eigenvalue* λ. Given a Markov matrix, M, a *stable* vector of M is an eigenvector with eigenvalue 1, i.e., $Mv = v$. A *distribution* is a vector $v \in \mathbb{R}^n$ such that $v \geq 0$ and $\|v\|_1 = Jv = 1$. So, a *stable distribution* is a stable vector that is also a distribution.

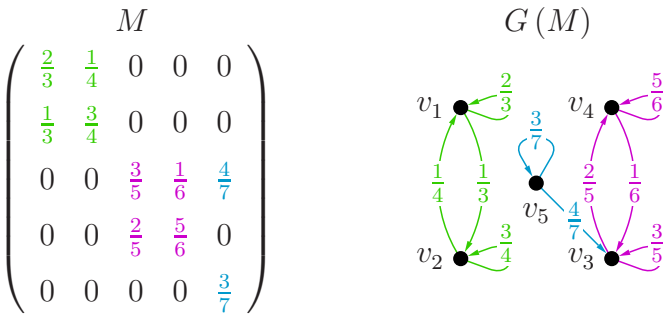
Observe that the set of stable vectors of M is a subset of the kernel[4] of Λ, since $Mv = v = Iv$ implies that $(M-I)v = 0$ so that $\Lambda v = 0$. More specifically, the stable distributions of M are precisely the non-negative, norm-1 vectors in $\ker \Lambda$, i.e., $\mathrm{stab}\,(M) = \ker \Lambda \cap \Delta_n$. Here, $\Delta_n = \{x \geq 0 \mid \forall i, x_i \geq 0 \text{ and } \sum_{i=1}^n x_i = 1\}$, the standard n-simplex.

We naturally associate a weighted graph $G(M) = (V, E, d)$ with any non-negative matrix, $M \geq 0$. Specifically, let $V = \{v_1, \ldots, v_n\}$, with $(v_i, v_j) \in E$ iff $M_{j,i} > 0$ and $d(v_i, v_j) = M_{j,i}$. Graphs obtained in this way cannot have multiple edges with the same starting and ending vertices. Also, every vertex in such a graph must have at least one outgoing edge. By ignoring the weights on $G(M)$, we obtain the corresponding unweighted graph, $G_{\_}(M) = (V, E)$.

For our purposes, M will usually be a Markov or a sub-Markov matrix. For example, the Markov matrix M on the left of Figure 1.1 gives rise to the "Markov" weighted graph

[4]Basic linear algebra concepts that are not defined in the main body of the thesis are reviewed in Appendix A.

Figure 1.1: Markov Matrix and its associated "Markov" Weighted Graph



on the right. Intuitively, the entries of M correspond to probabilities of traversing the corresponding edges. We do not include an edge from j to i in the graph when $M_{j,i} = 0$, since there is 0 probability of traversing such an edge, so it should not be the case that $v_i \rightsquigarrow v_j$, i.e., there should not be a walk (or path) from j to i.

A Markov matrix M is said to be *reducible* if $G(M)$ consists of more than one SCC; otherwise it is said to be *irreducible*. To conform with the literature on Markov chains, we call a Markov matrix *unichain* iff it has exactly one closed class. By Lemma 1.1, we can be sure that every Markov matrix has at least one closed class. Further, by Lemma 1.5, if we increase the number of non-zero entries of M, the number of closed classes cannot increase and must eventually decrease, since a complete graph consists of exactly one SCC, which is necessarily closed.

We will carry over the terminology of strongly connected components, closed classes, and invariant and transient sets of vertices in $G(M)$ and apply it to subsets of the indices of M in S_n. For example, $s \subset S_n$ is closed iff $V_s = \{v_i \mid i \in s\}$ is closed in $G(M)$. We can also define the submatrix, $M_{s,s'}$, of M corresponding to two subsets of indices $s, s' \subset S_n$ by removing row i and column j from M iff $i \notin s$ and $j \notin s'$. This submatrix is principal iff $s = s'$, in which case we say that $M_{s,s}$ is the principal submatrix of M corresponding to s. In the next chapter, we will present a more explicit means of constructing such submatrices.

Chapter 2

Existence of a Stable Vector

Because the reduction construction we present in Chapter 5 is defined in terms of submatrices, in this chapter, we carefully lay the groundwork for proving theorems about submatrices. Given $s \subset S_n$, we define two special matrices, π_s and $\imath_s$, that we use to extract the rows and columns, respectively, whose indices are in s of another matrix M. We then demonstrate how π_s and $\imath_s$ can be used to permute a matrix, yielding a partition that isolates the submatrix, $M_{s,s}$. Further, we prove that $\imath_s$ is always injective and that π_s is always surjective, and we show how their corresponding images and kernels are intimately related. Finally, and most notably,we show that the laplacian of any sub-Markov matrix, corresponding to a set s', has a non-zero kernel, if s' contains a closed class. In particular, the laplacian of any Markov matrix has a non-zero kernel. While this does *not* prove the existence of a stable distribution (because the stable vector need not be non-negative), it does hint at this important fact, which we will prove in Chapter 3.

2.1 Submatrix Construction

Given a subset of indices, $s \subset S_n$, with cardinality $k = |s|$, we can uniquely enumerate s in increasing order to obtain a sequence $(s_i)_{i=1}^k$. Mathematically, such a sequence is a bijective mapping from S_k to s, so we can also define its inverse, $s^{-1} : s \to S_k$, such that for $j \in s$, $s^{-1}(j) = i$ iff $s_i = j$. Further, we can enumerate the complement of s, $\overline{s}$, which has cardinality $\overline{k} = n - k$, and its inverse in exactly the same way. For example, if $s = \{1, 4\} \subset S_4$, so that $\overline{s} = \{2, 3\} \subset S_4$, then $s_1 = 1$, $s_2 = 4$, $s^{-1}(1) = 1$, and

$s^{-1}(4) = 2$; and $\overline{s}_1 = 2$, $\overline{s}_2 = 3$, $\overline{s}^{-1}(2) = 1$, and $\overline{s}^{-1}(3) = 2$.

Equipped with this notation, we can now present our method for constructing submatrices. For $s \subset S_n$, we will define the matrix

$$\imath_s = \begin{pmatrix} e_{s_1} & \cdots & e_{s_k} \end{pmatrix}.$$

It is easy to check that multiplying an $n \times n$ matrix, M, on the right by $\imath_s$ eliminates the columns of M whose indices are not in s and leaves the other columns intact, meaning in the same order. We will also define the matrix

$$\pi_s = \imath_s^t = \begin{pmatrix} e_{s_1}^t \\ \vdots \\ e_{s_k}^t \end{pmatrix}.$$

Again, it is easy to check that multiplying an $n \times n$ matrix, M, on the left by π_s eliminates the rows of M whose indices are not in s and leaves the other rows intact.

Now, given an $n \times n$ matrix M and two subsets $s, s' \subset S_n$, $M_{s,s'} \equiv \pi_s M \imath_{s'}$ is the submatrix that results from removing row i and column j from M iff $i \notin s$ and $j \notin s'$. Notice that $\left(M_{s,s'}\right)_{i,j} = e_i^t \pi_s M \imath_{s'} e_j = \left(\imath_s e_i\right)^t M \imath_{s'} e_j = e_{s_i}^t M e_{s'_j} = M_{s_i,s'_j}$.

Example 2.1. For example, let $s = \{1, 4\}$, and $M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}$. Here, $\imath_s = \begin{pmatrix} e_1 & e_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$, and $\pi_s = \begin{pmatrix} e_1^t \\ e_4^t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Further, $M\imath_s = \begin{pmatrix} 1 & 4 \\ 5 & 8 \\ 9 & 12 \\ 13 & 16 \end{pmatrix}$, $\pi_s M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 13 & 14 & 15 & 16 \end{pmatrix}$, and $\pi_s M \imath_s = \begin{pmatrix} 1 & 4 \\ 13 & 16 \end{pmatrix} = M_{s,s}$. Hence, $M_{s,s}$ is the (principal) submatrix corresponding to s. □

2.2 Matrix Permutations

A *permutation* of a set s is a bijective mapping $s \to s$. When $s \subseteq S_n$ is viewed as a sequence, we can think of a permutation as a reordering of the elements of s.

Given a permutation ρ of S_n, the *matrix permutation* of an $n \times n$ matrix M according to ρ is a rearrangement of M's entries resulting in the permuted matrix M' such that $M_{i,j} = M'_{\rho(i),\rho(j)}$, or equivalently, $M'_{i,j} = M_{\rho^{-1}(i),\rho^{-1}(j)}$. In the graph-theoretic representation of M, $G(M)$, this corresponds to relabelling the vertices, without changing the edges or their weights.

Using the submatrix construction given above, for any subset $s \subset S_n$, we can define permutation matrices P_s and P_s^t such that $P_s^t M P_s$ is a permutation of M that moves the principal submatrix $M_{s,s}$ to the lower-right-hand corner of M.

If $P_s = \begin{pmatrix} \imath_{\overline{s}} & \imath_s \end{pmatrix}$, so that $P_s^t = \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix}$, then

$$P_s^t M P_s = \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix} M \begin{pmatrix} \imath_{\overline{s}} & \imath_s \end{pmatrix} = \begin{pmatrix} \pi_{\overline{s}} M \imath_{\overline{s}} & \pi_{\overline{s}} M \imath_s \\ \pi_s M \imath_{\overline{s}} & \pi_s M \imath_s \end{pmatrix} = \begin{pmatrix} M_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & M_{s,s} \end{pmatrix}.$$

We will refer to this collection of sub-matrices $M_{s,s}$, $M_{\overline{s},s}$, $M_{s,\overline{s}}$, and $M_{\overline{s},\overline{s}}$ as a *partitioning* of M with respect to s. Now recalling our enumeration of s and $\overline{s}$, with cardinalities k and $\overline{k}$, respectively, let

$$\rho_s(i) = \begin{cases} \overline{s}^{-1}(i),\ i \in \overline{s} \\ \overline{k} + s^{-1}(i),\ i \in s \end{cases}.$$

We will see that $P_s^t M P_s$ is a matrix permutation of M according to ρ_s.

Example 2.2. With $s = \{1, 4\} \subset S_4$ and $M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}$ as above,

$$P_s^t M P_s = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 7 & 5 & 8 \\ 10 & 11 & 9 & 12 \\ 2 & 3 & 1 & 4 \\ 14 & 15 & 13 & 16 \end{pmatrix},$$

so that M has the partitioning $M_{s,s} = \begin{pmatrix} 1 & 4 \\ 13 & 16 \end{pmatrix}$, $M_{\overline{s},s} = \begin{pmatrix} 5 & 8 \\ 9 & 12 \end{pmatrix}$, $M_{s,\overline{s}} = \begin{pmatrix} 2 & 3 \\ 14 & 15 \end{pmatrix}$, and $M_{\overline{s},\overline{s}} = \begin{pmatrix} 6 & 7 \\ 10 & 11 \end{pmatrix}$ with respect to s.

Notice that $M_{i,j} = (P_s^t M P_s)_{\rho_s(i),\rho_s(j)}$. For example, $M_{1,2} = 2$, and $(P_s^t M P_s)_{\rho_s(1),\rho_s(2)} = (P_s^t M P_s)_{3,1} = 2$. □

Theorem 2.3. *Given $s \subset S_n$, with $\rho_s : S_n \to S_n$ and the matrices P_s, P_s^t as defined above, ρ_s is a permutation of S_n, and $P_s^t M P_s$ is the permutation of M according to ρ_s for any $n \times n$ matrix M.*

Proof. First we will show that ρ_s is a permutation of S_n. Recall that s^{-1} and $\overline{s}^{-1}$ are bijective mappings $s \to S_k$ and $\overline{s} \to S_{\overline{k}}$, with $k + \overline{k} = n$. So for $i \in S_n$, ρ_s is a bijective mapping $\overline{s} \to S_{\overline{k}}$ if $i \in \overline{s}$, and ρ_s is a bijective mapping $s \to (\overline{k} + S_k)$ if $i \in s$. Here $(\overline{k} + S_k) \equiv \{\overline{k} + l \mid l \in S_k\}$. In the first case, $1 \leq \rho_s(i) \leq \overline{k}$, and in the second, $\overline{k} + 1 \leq \rho_s(i) \leq \overline{k} + k = n$. The domains $s, \overline{s}$ are disjoint with $s \cup \overline{s} = S_n$, and the images $S_{\overline{k}}, (\overline{k} + S_k)$ are disjoint with $S_{\overline{k}} \cup (\overline{k} + S_k) = S_n$, so ρ_s is a bijective mapping $S_n \to S_n$, that is, a permutation of S_n. This leads to the formula for

$$\rho_s^{-1}(i) = \begin{cases} \overline{s}_i, \ 1 \leq i \leq \overline{k} \\ s_{i-\overline{k}}, \ \overline{k} + 1 \leq i \leq n \end{cases}.$$

Now for $M' = P_s^t M P_s$, we will show that $M'_{i,j} = M_{\rho_s^{-1}(i),\rho_s^{-1}(j)}$. Notice that $M'_{i,j} = e_i^t M' e_j = e_i^t P_s^t M P_s e_j = (P_s e_i)^t M (P_s e_j)$. But $P_s e_i$ is the i^{th} column of P_s, which by definition is $e_{\overline{s}_i}$ if $i \leq \overline{k}$ and $e_{s_{i-\overline{k}}}$ otherwise. In other words, $P_s e_i = e_{\rho_s^{-1}(i)}$. Similarly, $P_s e_j = e_{\rho_s^{-1}(j)}$. So $M'_{i,j} = e^t_{\rho_s^{-1}(i)} M e_{\rho_s^{-1}(j)} = M_{\rho_s^{-1}(i),\rho_s^{-1}(j)}$, and $P_s^t M P_s$ is the permutation of M according to ρ_s. □

Corollary 2.4. *For the permutation matrix P_s corresponding to $s \subset S_n$, $P_s^t = P_s^{-1}$, that is, $P_s^t P_s = I$.*

Proof. $P_s^t P_s = P_s^t I P_s = I'$, the permutation of I according to ρ_s. Then $I'_{i,j} = I_{\rho_s^{-1}(i),\rho_s^{-1}(j)}$, and since ρ_s^{-1} is bijective, $\rho_s^{-1}(i) = \rho_s^{-1}(j)$ iff $i = j$. So $I'_{i,j} = 1$ iff $i = j$, and $I'_{i,j} = 0$ otherwise. That is, $I' = I$, so $P_s^t = P_s^{-1}$. □

Corollary 2.5. *For the permutation matrix P_s corresponding to $s \subset S_n$, P_s and P_s^t are Markov. The product of Markov matrices (of the same dimension) is Markov, and, in particular, and $P_s^t M P_s$ is Markov for any $n \times n$ Markov matrix, M.*

Proof. By definition, the columns of P_s are the standard basis vectors. Thus they are non-negative and sum to 1, and P_s is Markov. In particular, $JP_s = J$. Therefore, $J = JP_sP_s^{-1} = JP_s^{-1} = JP_s^t$, so that the columns of P_s^t sum to 1 as well. Since P_s^t is non-negative, P_s^t is also Markov.

If M_1, M_2 are Markov and of the same dimension, then $M_1M_2 \geq 0$ and $J(M_1M_2) = (JM_1)M_2 = JM_2 = J$, so that M_1M_2 is Markov. Since P_s^t and P_s are Markov, so is $P_s^t M P_s$, when M is $n \times n$ and Markov. □

2.3 Projection and Inclusion

For a subset $s \subset S_n$, with $k = |s|$, $\imath_s$ has dimension $n \times k$, and π_s has dimension $k \times n$. So by left-multiplication, $\imath_s$ is a mapping $\mathbb{R}^k \to \mathbb{R}^n$, and π_s is a mapping $\mathbb{R}^n \to \mathbb{R}^k$. We call $\imath$ an *inclusion* operator, because for $v \in \mathbb{R}^k$, $w = \imath_s v \in \mathbb{R}^n$ is the vector whose coordinates with indices in s are given by the coordinates of v (in order), while its coordinates with indices in $\overline{s}$ are all 0. Likewise, we call π_s a *projection* operator, because for $v' \in \mathbb{R}^n$, the coordinates of $w' = \pi_s v' \in \mathbb{R}^k$ are just the coordinates of v' with indices in s.

Example 2.6. Let $s = \{1, 4\} \subset S_4$ and $v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4$. The vector $w = \pi_s v = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ is the projection of v on $\mathbb{R}^2$. The vector $u = \imath_s w =$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 4 \end{pmatrix}$$ is the inclusion of w in $\mathbb{R}^4$. □

Next, we will define $\mathbb{R}^s \equiv \operatorname{span}\{e_i \mid i \in s\}$, a subspace of $\mathbb{R}^n$ of dimension k. Similarly, $\mathbb{R}^{\overline{s}} \equiv \operatorname{span}\{e_i \mid i \in \overline{s}\}$ is a subspace of $\mathbb{R}^n$ of dimension $\overline{k}$. The next lemma highlights the key algebraic and geometric properties of these projection and inclusion operators.

Lemma 2.7. *Given $s \subset S_n$,*

a) $\pi_s \imath_s = I = \pi_{\overline{s}} \imath_{\overline{s}}$, $\pi_{\overline{s}} \imath_s = 0 = \pi_s \imath_{\overline{s}}$, *and* $\imath_{\overline{s}} \pi_{\overline{s}} + \imath_s \pi_s = I$;

b) $\ker \imath_{\overline{s}} = 0 = \ker \imath_s$, $\operatorname{im} \pi_{\overline{s}} = \mathbb{R}^{\overline{k}}$, *and* $\operatorname{im} \pi_s = \mathbb{R}^k$;

c) $\operatorname{im} \imath_{\overline{s}} = \mathbb{R}^{\overline{s}} = \ker \pi_s$, *and* $\operatorname{im} \imath_s = \mathbb{R}^s = \ker \pi_{\overline{s}}$.

Proof. Proof of part a): Since P_s and P_s^t are inverses,

$$I = P_s^t P_s = \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix} \begin{pmatrix} \imath_{\overline{s}} & \imath_s \end{pmatrix} = \begin{pmatrix} \pi_{\overline{s}} \imath_{\overline{s}} & \pi_{\overline{s}} \imath_s \\ \pi_s \imath_{\overline{s}} & \pi_s \imath_s \end{pmatrix}$$

so that $\pi_{\overline{s}} \imath_{\overline{s}} = \pi_s \imath_s = I$ and $\pi_{\overline{s}} \imath_s = \pi_s \imath_{\overline{s}} = 0$. Likewise,

$$I = P_s P_s^t = \begin{pmatrix} \imath_{\overline{s}} & \imath_s \end{pmatrix} \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix} = \imath_{\overline{s}} \pi_{\overline{s}} + \imath_s \pi_s$$

Proof of part b): By part a), $\pi_{\overline{s}} \imath_{\overline{s}} = \pi_s \imath_s = I$. So, $\imath_{\overline{s}}$ and $\imath_s$ are left-invertible, hence injective with $\ker \imath_{\overline{s}} = \ker \imath_s = \{0\}$. Likewise, $\pi_{\overline{s}}$ and π_s are right-invertible, hence surjective with $\operatorname{im} \pi_{\overline{s}} = \mathbb{R}^{\overline{k}}$ and $\operatorname{im} \pi_s = \mathbb{R}^k$.

Proof of part c): The fact that $\operatorname{im} \imath_{\overline{s}} = \mathbb{R}^{\overline{s}}$ can be seen as follows:

$$\begin{aligned}
\operatorname{im} \imath_s &= \left\{ \imath_s v \mid v \in \mathbb{R}^k \right\} \\
&= \left\{ \imath_s \sum_{i=1}^{k} v_i e_i \mid v_i \in \mathbb{R}, e_i \in \mathbb{R}^k \right\} \\
&= \left\{ \sum_{i=1}^{k} v_i \imath_s e_i \mid v_i \in \mathbb{R}, e_i \in \mathbb{R}^k \right\} \\
&= \left\{ \sum_{i=1}^{k} v_i e_{s_i} \mid v_i \in \mathbb{R}, e_{s_i} \in \mathbb{R}^n \right\} \\
&= \operatorname{span} \left\{ e_j \mid j \in s \right\} \\
&= \mathbb{R}^s
\end{aligned}$$

Likewise, $\operatorname{im} \imath_{\overline{s}} = \mathbb{R}^{\overline{s}}$.

Now since $\pi_{\overline{s}} \imath_s = 0$, it follows that $\operatorname{im} \imath_s \subset \ker \pi_{\overline{s}}$. Conversely, if $v \in \ker \pi_{\overline{s}}$, then by part a), $v = \left(\imath_{\overline{s}} \pi_{\overline{s}} + \imath_s \pi_s \right) v = \imath_{\overline{s}} \pi_{\overline{s}} v + \imath_s \pi_s v = \imath_s \pi_s v$. But $\imath_s \pi_s v \in \operatorname{im} \imath_s$. Thus, $\ker \pi_{\overline{s}} \subset \operatorname{im} \imath_s$, so that $\ker \pi_{\overline{s}} = \operatorname{im} \imath_s = \mathbb{R}^s$. Likewise, $\operatorname{im} \imath_{\overline{s}} = \ker \pi_s = \mathbb{R}^{\overline{s}}$.

□

The compound operation $\imath_s \pi_s$ takes any vector $v \in \mathbb{R}^n$ and maps it to a vector $u \in \mathbb{R}^s$ such that $u_i = v_i$ if $i \in s$ and $u_i = 0$ otherwise. This allows us to take any vector, $v \in \mathbb{R}^n$, and easily decompose it as $v = v_s + v_{\overline{s}}$, with $v_s \in \mathbb{R}^s$ and $v_{\overline{s}} \in \mathbb{R}^{\overline{s}}$.

Example 2.8. As in Example 2.6, let $s = \{1, 4\} \subset S_4$ and $v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4$. Define $v_s = \imath_s \pi_s v = \imath_{\overline{s}} \pi_{\overline{s}} v = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 4 \end{pmatrix}$. Similarly, define

$v_{\overline{s}} = \imath_{\overline{s}}\pi_{\overline{s}}v = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 3 \\ 0 \end{pmatrix}$. Now observe that $v_s \in \mathbb{R}^s$, $v_{\overline{s}} \in \mathbb{R}^{\overline{s}}$, and $v_s + v_{\overline{s}} = v$. □

Theorem 2.9. *Given $s \subset S_n$, $\mathbb{R}^n = \mathbb{R}^{\overline{s}} \oplus \mathbb{R}^s$. Further, for any $v \in \mathbb{R}^n$, $v = v_{\overline{s}} + v_s$, with $v_{\overline{s}} = \imath_{\overline{s}}\pi_{\overline{s}}v \in \mathbb{R}^{\overline{s}}$ and $v_s = \imath_s\pi_s v \in \mathbb{R}^s$.*

Proof. Suppose $v \in \mathbb{R}^{\overline{s}} \cap \mathbb{R}^s$, $v \neq 0$. Then $v = \sum_{i\in\overline{s}} a_i e_i$, and $v = \sum_{j\in s} b_j e_j$, with $a_i, b_j \neq 0$ for some i, j. But then $\sum_{i\in\overline{s}} a_i e_i - \sum_{j\in s} b_j e_j = 0$, which is impossible since the standard basis vectors are linearly independent. So $\mathbb{R}^{\overline{s}} \cap \mathbb{R}^s = 0$.

Now for any $v \in \mathbb{R}^n$, by Lemma 2.7 a), $v = Iv = \left(\imath_{\overline{s}}\pi_{\overline{s}} + \imath_s\pi_s\right) v = \imath_{\overline{s}}\pi_{\overline{s}}v + \imath_s\pi_s v = v_{\overline{s}} + v_s$. Since $\operatorname{im}\imath_{\overline{s}} = \mathbb{R}^{\overline{s}}$ and $\operatorname{im}\imath_s = \mathbb{R}^s$, it follows that $v_{\overline{s}} \in \mathbb{R}^{\overline{s}}$ and $v_s \in \mathbb{R}^s$. But then $v \in \mathbb{R}^{\overline{s}} + \mathbb{R}^s$, so that $\mathbb{R}^n = \mathbb{R}^{\overline{s}} + \mathbb{R}^s$. Therefore, $\mathbb{R}^n = \mathbb{R}^{\overline{s}} \oplus \mathbb{R}^s$. □

2.4 Existence Theorem

In this section, we compile a collection of basic facts regarding the structure of a Markov matrix, its closed classes, and stable vectors, which will be needed in subsequent chapters. Most notably, we show that the laplacian of any sub-Markov matrix, corresponding to a set s', has a non-zero kernel, if s' contains a closed class. In particular, the laplacian of any Markov matrix has a non-zero kernel, which contains a linearly independent set of vectors corresponding to its closed classes.

Lemma 2.10. *Let M be an $n \times n$ Markov matrix, and a subset of indices $s \subset S_n$, where s is a closed class or union of closed classes of M.*

a) $M_{\overline{s},s} = 0$.

b) $M_{s,s}$ *is Markov.*

c) If s' is a subset of indices such that $s \subset s' \subset S_n$, then $\imath_s(\ker\Lambda_{s,s}) \subset \imath_{s'}(\ker\Lambda_{s',s'})$.

d) $\imath_s\left(\operatorname{stab} M_{s,s}\right) \subset \operatorname{stab} M$.

Proof. Proof of part a): By assumption, s is a union of closed classes. So there are no edges in $G(M)$ from s to $\overline{s}$, which means that $M_{\overline{s}_i,s_j} = 0$, for any $1 \leq i \leq |\overline{s}|$ and $1 \leq j \leq |s|$. But $\left(M_{\overline{s},s}\right)_{i,j} = M_{\overline{s}_i,s_j}$, so $\left(M_{\overline{s},s}\right)_{i,j} = 0$, and $M_{\overline{s},s}$ is the zero matrix.

Proof of part b): By part a),

$$P_s^t M P_s = \begin{pmatrix} M_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & M_{s,s} \end{pmatrix} = \begin{pmatrix} M_{\overline{s},\overline{s}} & 0 \\ M_{s,\overline{s}} & M_{s,s} \end{pmatrix}.$$

Since $P_s^t M P_s$ is Markov, its columns sum to 1, and in particular the columns of $M_{s,s}$ sum to 1. Further $M_{s,s} \geq 0$, since $M \geq 0$. Hence, $M_{s,s}$ is Markov.

Proof of part c): Take any $v \in \ker \Lambda_{s,s}$, so that $M_{s,s}v = v$. Because $s \subset s'$, $\mathbb{R}^s \subset \mathbb{R}^{s'}$, and so $\operatorname{im} \imath_s \subset \operatorname{im} \imath_{s'}$, since $\operatorname{im} \imath_s = \mathbb{R}^s$ and $\operatorname{im} \imath_{s'} = \mathbb{R}^{s'}$. In particular, $\imath_s v \in \operatorname{im} \imath_{s'}$, so there exists a $v' \in \mathbb{R}^{|s'|}$ such that $\imath_{s'} v' = \imath_s v$. Now observe the following:

$$\begin{aligned}
M_{s',s'}v' &= \pi_{s'} M \imath_{s'} v' \\
&= \pi_{s'} M \imath_s v \\
&= \pi_{s'} I M \imath_s v \\
&= \pi_{s'} \left(\imath_{\overline{s}}\pi_{\overline{s}} + \imath_s \pi_s\right) M \imath_s v && \text{by Lemma 2.7 a)} \\
&= \pi_{s'} \imath_{\overline{s}} \pi_{\overline{s}} M \imath_s v + \pi_{s'} \imath_s \pi_s M \imath_s v \\
&= \pi_{s'} \imath_{\overline{s}} M_{\overline{s},s} v + \pi_{s'} \imath_s M_{s,s} v \\
&= 0 + \pi_{s'} \imath_s M_{s,s} v && \text{by part a)} \\
&= \pi_{s'} \imath_s v && \text{by assumption} \\
&= \pi_{s'} \imath_{s'} v' \\
&= v' && \text{by Lemma 2.7 a)}
\end{aligned}$$

Thus, $v' \in \ker \Lambda_{s',s'}$, and $\imath_s \left(\ker \Lambda_{s,s}\right) \subset \imath_{s'} \left(\ker \Lambda_{s',s'}\right)$.

Proof of part d): Consider $v \in \operatorname{stab} M_{s,s}$, so $v \geq 0$, $Jv = 1$, and $v \in \ker \Lambda_{s,s}$. We will apply part c) with $s' = S_n$. The columns of $\imath_{s'}$ are the standard basis vectors in $\mathbb{R}^n$, e_i, such that $i \in s'$. But since $s' = S_n$, this includes all of them, and $\imath_{s'} = I$. So $\tilde{v} = \imath_s v \in I \ker \Lambda = \ker \Lambda$. Hence, we need only show that $\tilde{v}$ is a distribution. Since

$v, \imath_s \geq 0$, it follows that $\tilde{v} \geq 0$. $\tilde{v}_i = v_{s^{-1}(i)}$ for $i \in s$, and $\tilde{v}_i = 0$ otherwise, so $J\tilde{v} = \sum_{i=1}^{n} \tilde{v}_i = \sum_{i \in s} \tilde{v}_i = \sum_{i=1}^{|s|} v_i = Jv = 1$. so that $\tilde{v} \in \operatorname{stab} M$.
□

Theorem 2.11. *Let M be an $n \times n$ Markov matrix.*

a) $\ker \Lambda \neq 0$.

b) *For any subset of indices $s' \subset S_n$ such that s' contains a closed class,* $\ker \Lambda_{s',s'} \neq 0$.

Proof. Proof of part a): The matrix M is Markov, so $JM = J$. But then $M^t J^t = (JM)^t = J^t$. This implies that $(M^t - I)J^t = 0$, so $J^t \in \ker \Lambda^t$, meaning $\dim \ker \Lambda^t \neq 0$. Finally, by Theorem A.1, $\dim \ker \Lambda \neq 0$, so $\ker \Lambda \neq 0$.

Proof of part b): By assumption, s' contains a closed class. Call it s. By Lemma 2.10 b), $M_{s,s}$ is Markov, so by part a), $\ker \Lambda_{s,s} \neq 0$. Pick $v \in \ker \Lambda_{s,s}$ such that $v \neq 0$. Now by Lemma 2.10 c), $\imath_s(\ker \Lambda_{s,s}) \subset \imath_{s'}\left(\ker \Lambda_{s',s'}\right)$, so there exists $v' \in \ker \Lambda_{s',s'}$ such that $\imath_s v = \imath_{s'} v'$. Further, by Lemma 2.7 b), $\imath_s$ is injective, so $v' \neq 0$. Therefore, $\ker \Lambda_{s',s'} \neq 0$.
□

As a consequence of this theorem, every Markov matrix M has a stable vector. In fact, this is true of any principal submatrix, $M_{s,s}$, of M containing a closed class, s. By Lemma 2.10 c), any stable vector of $M_{s,s}$ can be extended to be a stable vector of M. Thus, the kernel of Λ contains a stable vector of M corresponding to each of its closed classes, and these vectors are necessarily independent (since they are non-zero on disjoint sets of indices). Once we show, in Chapter 3, that every Markov matrix has a *non-negative* stable vector, and hence a stable distribution, Lemma 2.10 d) will likewise guarantee the existence of a set of independent stable *distributions* corresponding to the closed classes of M.

Chapter 3

Existence of a Stable Distribution

Given any distribution, v_0, and a Markov matrix, M, of the same dimension, we can construct a sequence of distributions via iteration, $v_i = Mv_{i-1}$, $i \geq 1$. While v_i need not converge as $i \to \infty$, it necessarily converges *in the Cesaro sense* (Marsden, 1974, p. 363), that is, $\frac{1}{N}\sum_{i=0}^{N-1} v_i$ converges as $N \to \infty$. More generally, for any Markov matrix, M^j converges in the Cesaro sense (Doob, 1953), that is, the matrix, $M^\infty \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{j=0}^{N-1} M^j$ is well-defined. Since $v_i = M^i v_0$, this implies that v_i converges in the Cesaro sense to $\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} v_i = \lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} M^i v_0 = M^\infty v_0$.

In this chapter, we prove a sharper result. We show that the laplacian, Λ, of M induces a natural splitting of $\mathbb{R}^n$ into the kernel and the image of Λ, and that M^∞ is the associated projection, $\pi_{\ker}$, onto $\ker \Lambda$. This allows us to prove our first structure theorem for Markov matrices, in which we characterize the set of stable distributions of a Markov matrix, M, in terms of the columns of M^∞. Specifically, we prove that $M^\infty e_i \in \mathrm{stab}(M)$, for all $1 \leq i \leq n$. This proves the existence of a stable distribution for any Markov matrix.

Lemma 3.1. *If M is Markov, then*

a) $\ker \Lambda \oplus \mathrm{im}\, \Lambda = \mathbb{R}^n$,

b) $\alpha : \ker \Lambda \times \mathrm{im}\, \Lambda \to \mathbb{R}^n$, *such that* $\alpha(v, w) = v + w$, *is linear and invertible,*

c) *there are well-defined linear mappings,* $\pi_{\ker} : \mathbb{R}^n \to \ker \Lambda$ *and* $\pi_{\mathrm{im}} : \mathbb{R}^n \to \mathrm{im}\, \Lambda$ *such that*

i) $\alpha(\pi_{\ker}, \pi_{\mathrm{im}}) = \pi_{\ker} + \pi_{\mathrm{im}} = I$,

ii) $\pi_{\ker}|_{\ker\Lambda} = I$, $\pi_{\mathrm{im}}|_{\mathrm{im}\,\Lambda} = I$,

iii) $\mathrm{im}\,\pi_{\ker} = \ker\Lambda = \ker\pi_{\mathrm{im}}$, *and*

iv) $\mathrm{im}\,\pi_{\mathrm{im}} = \mathrm{im}\,\Lambda = \ker\pi_{\ker}$.

Proof. We first show that $\ker\Lambda \cap \mathrm{im}\,\Lambda = 0$. Take $v \in \ker\Lambda \cap \mathrm{im}\,\Lambda$. Since $v \in \mathrm{im}\,\Lambda$, there exists w such that $v = \Lambda w = (M - I)w = Mw - w$ so that $Mw = v + w$. In addition, since $v \in \ker\Lambda$, $\Lambda v = 0$, and $Mv = v$. Therefore, by a straightforward induction, $M^k w = w + kv$ for any $k \geq 1$. Solving for v yields $v = \frac{1}{k}\left(M^k - I\right)w$. But $\|v\|_1 = \left\|\frac{1}{k}\left(M^k - I\right)w\right\|_1 \leq \frac{1}{k}\left\|M^k - I\right\|_1 \|w\|_1 \leq \frac{1}{k}\left(\left\|M^k\right\|_1 + \|I\|_1\right)\|w\|_1 \leq \frac{1}{k}\left(\|M\|_1^k + \|I\|_1\right)\|w\|_1 = \frac{2}{k}\|w\|_1$. Since $0 \leq \|v\|_1 \leq \frac{2}{k}\|w\|_1$, which is as small as we like for large k, it follows that $\|v\|_1 = 0$, and therefore $v = 0$.

For general vector spaces, $\dim(V + W) = \dim V + \dim W - \dim(V \cap W)$. Applying this identity to $V = \ker\Lambda$ and $W = \mathrm{im}\,\Lambda$, and using the fact that $V \cap W = 0$, we have that $\dim(\ker\Lambda + \mathrm{im}\,\Lambda) = \dim\ker\Lambda + \dim\mathrm{im}\,\Lambda$. By Theorem A.1, $\dim\mathrm{im}\,M + \dim\ker M = n$ for any $m \times n$-dimensional matrix, M, $\dim(\ker\Lambda + \mathrm{im}\,\Lambda) = n$, and $\ker\Lambda + \mathrm{im}\,\Lambda \subset \mathbb{R}^n$, so it must be the case that $\ker\Lambda \oplus \mathrm{im}\,\Lambda = \mathbb{R}^n$.

This means that the mapping $\alpha : \ker\Lambda \times \mathrm{im}\,\Lambda \to \mathbb{R}^n$, such that $\alpha(v, w) = v + w$, for $v \in \ker\Lambda$ and $w \in \mathrm{im}\,\Lambda$, is surjective. It is also injective. If $0 = \alpha(v, w) = v + w$, then $v = -w$, that is, v and w are multiples of one another. But $v \in \ker\Lambda$ and $w \in \mathrm{im}\,\Lambda$, so $v, w \in \ker\Lambda \cap \mathrm{im}\,\Lambda = 0$. In particular, $v = w = 0$, so $\ker\alpha = (0, 0)$. Further, α is linear, since it is just addition.

Thus, there exists an inverse linear mapping, $\alpha^{-1} : \mathbb{R}^n \to \ker\Lambda \times \mathrm{im}\,\Lambda$, corresponding to a pair of linear mappings, $\pi_{\ker}$ and π_{im}, with $\mathrm{im}\,\pi_{\ker} \subset \ker\Lambda$ and $\mathrm{im}\,\pi_{\mathrm{im}} \subset \mathrm{im}\,\Lambda$. In particular, $\alpha(\pi_{\ker}, \pi_{\mathrm{im}}) = \pi_{\ker} + \pi_{\mathrm{im}} = I$, since $v = \left(\alpha\alpha^{-1}\right)v = \alpha\left(\alpha^{-1}v\right) = \alpha\left(\pi_{\ker}v, \pi_{\mathrm{im}}v\right) = \pi_{\ker}v + \pi_{\mathrm{im}}v = \left(\pi_{\ker} + \pi_{\mathrm{im}}\right)v$.

Moreover, if $v \in \ker\Lambda$, since $\alpha(v, 0) = v$ and $(v, 0) = \alpha^{-1}(v) = (\pi_{\ker}v, \pi_{\mathrm{im}}v)$, $\pi_{\ker}v = v$. In particular, $\pi_{\ker}|_{\ker\Lambda} = I$, so that $\ker\Lambda \subset \mathrm{im}\,\pi_{\ker}$. Since by definition, $\mathrm{im}\,\pi_{\ker} \subset \ker\Lambda$, $\mathrm{im}\,\pi_{\ker} = \ker\Lambda$. Also, $\pi_{\mathrm{im}}v = 0$, so $\ker\Lambda \subset \ker\pi_{\mathrm{im}}$. Suppose $w \in \ker\pi_{\mathrm{im}}$. Then $w = Iw = (\pi_{\ker} + \pi_{\mathrm{im}})w = \pi_{\ker}w + \pi_{\mathrm{im}}w = \pi_{\ker}w$. Then $w \in \mathrm{im}\,\pi_{\ker} = \ker\Lambda$, and $\ker\pi_{\mathrm{im}} \subset \ker\Lambda$, so $\ker\pi_{\mathrm{im}} = \ker\Lambda$.

Similarly, if $v \in \mathrm{im}\,\Lambda$, since $\alpha(0, v) = v$ and $(0, v) = \alpha^{-1}(v) = (\pi_{\ker}v, \pi_{\mathrm{im}}v)$, $\pi_{\mathrm{im}}v = v$. In particular, $\pi_{\mathrm{im}}|_{\mathrm{im}\,\Lambda} = I$, so that $\mathrm{im}\,\Lambda \subset \mathrm{im}\,\pi_{\mathrm{im}}$. Since by definition, $\mathrm{im}\,\pi_{\mathrm{im}} \subset \mathrm{im}\,\Lambda$,

$\operatorname{im}\pi_{\text{im}} = \operatorname{im}\Lambda$. Also, $\pi_{\text{ker}}v = 0$, so $\operatorname{im}\Lambda \subset \ker\pi_{\text{ker}}$. Suppose $w \in \ker\pi_{\text{ker}}$. Then $w = Iw = (\pi_{\text{ker}} + \pi_{\text{im}})w = \pi_{\text{ker}}w + \pi_{\text{im}}w = \pi_{\text{im}}w$. Then $w \in \operatorname{im}\pi_{\text{im}} = \operatorname{im}\Lambda$, and $\ker\pi_{\text{ker}} \subset \operatorname{im}\Lambda$, so $\ker\pi_{\text{ker}} = \operatorname{im}\Lambda$.

□

Although we will prove that M^∞ is well-defined without it, the following lemma, which characterizes π_{ker}, will also will turn out to provide an interesting characterization of M^∞.

Lemma 3.2. *For any Markov matrix, M, there exists exactly one matrix, M', satisfying the following conditions:*

$$MM' = M' \tag{3.1}$$

$$M'M = M' \tag{3.2}$$

$$M'M' = M' \tag{3.3}$$

$$\operatorname{rk} M' = \dim\ker\Lambda \tag{3.4}$$

In fact, we must have $M' = \pi_{\text{ker}}$ and $I - M' = \pi_{\text{im}}$ from Lemma 3.1.

Proof. Existence. We begin by showing that $M' = \pi_{\text{ker}}$ satisfies Equations 3.1-3.4. Equation 3.1 is equivalent to $(M - I)\pi_{\text{ker}} = \Lambda\pi_{\text{ker}} = 0$, which is clearly true, since $\operatorname{im}\pi_{\text{ker}} = \ker\Lambda$. Likewise, Equation 3.2 is equivalent to $\pi_{\text{ker}}(M - I) = \pi_{\text{ker}}\Lambda = 0$, which is also true, since $\operatorname{im}\Lambda = \ker\pi_{\text{ker}}$. Equation 3.3 follows from Lemma 3.1 c) ii). Since $\operatorname{im}\pi_{\text{ker}} = \ker\Lambda$ and $\pi_{\text{ker}}|_{\ker\Lambda} = I$, we have $\pi_{\text{ker}}\pi_{\text{ker}} = I\pi_{\text{ker}} = \pi_{\text{ker}}$. Finally, Equation 3.4 follows from the definition of rank (cf. Theorem A.1) and the fact that $\operatorname{im}\pi_{\text{ker}} = \ker\Lambda$.

Uniqueness. Equation 3.1 implies that $\Lambda M' = 0$, so that $\operatorname{im}M' \subset \ker\Lambda$, and, by Equation 3.4, we have $\dim\operatorname{im}M' = \operatorname{rk}M' = \dim\ker\Lambda$, so that $\operatorname{im}M' = \ker\Lambda$. Likewise, Equation 3.2 implies that $M'\Lambda = 0$, so that $\operatorname{im}\Lambda \subset \ker M'$. By Equation 3.4, $\dim\operatorname{im}\Lambda = n - \dim\ker\Lambda = n - \operatorname{rk}M' = n - \dim\operatorname{im}M' = \dim\ker M'$, so $\operatorname{im}\Lambda = \ker M'$ as well.

By Equation 3.3, $M'(I - M') = M' - M' = 0$, so that $\operatorname{im}(I - M') \subset \ker M'$. Thus, $\left(M', (I - M')\right)$ is mapping from $\mathbb{R}^n$ to $\operatorname{im}M' \times \operatorname{im}\Lambda' = \operatorname{im}M' \times \ker M' = \ker\Lambda \times \operatorname{im}\Lambda$ such that, using the notation of Lemma 3.1, $\alpha\left(M', \Lambda'\right) = M' + \Lambda' = I$. Since inverses are unique, we must have that $M' = \pi_{\text{ker}}$ and $I - M' = \pi_{\text{im}}$. □

Next, we prove that M^∞ exists, and further, that it is the unique matrix satisfying the conditions of Lemma 3.2. In particular, $M^\infty = \pi_{\text{ker}}$. Hence, π_{ker} is Markov.

Theorem 3.3. *For any Markov matrix, M, the sequence of Markov matrices, $M_N \equiv \frac{1}{N}\sum_{j=0}^{N-1} M^j$ converges to M^∞ as $N \to \infty$. Moreover, $M^\infty = \pi_{\text{ker}}$, and it is Markov.*

Proof. First, we show that M^∞, the limit of M_N as $N \to \infty$, is well-defined. We will appeal to the classic result from real analysis which says that a sequence converges iff every subsequence has a convergent subsequence with a common limit (Royden, 1968, p. 37, ex. 11). Since M_N is a bounded sequence ($0 \leq \left\|M^j\right\|_1 \leq 1$, so that $\sum_{j=0}^{N-1}\left\|M^j\right\|_1 \leq N$), any subsequence is also bounded. Thus, by a standard argument from real analysis, (Royden, 1968, p. 37, ex. 8), every subsequence, in turn, has a convergent subsequence, call it M_{N_i}.

By Lemma 3.1, any vector, $v = \pi_{\text{ker}}v + \pi_{\text{im}}v$. If $\{v_1, \ldots, v_k\}$ is a basis for $\ker\Lambda$, then there exist $\beta_r \in \mathbb{R}^k$ and $w \in \mathbb{R}^n$ such that $\pi_{\text{ker}}v = \sum_{r=1}^k \beta_r v_r \in \ker\Lambda$ (i.e., $\sum_{r=1}^k \beta_r v_r$ is a stable distribution of M^j, $j \geq 0$) and $\pi_{\text{im}}v = \Lambda w$. So,

$$\begin{aligned} M_{N_i}v &= \frac{1}{N_i}\sum_{j=0}^{N_i-1} M^j v \\ &= \frac{1}{N_i}\sum_{j=0}^{N_i-1} M^j\left[\sum_{r=1}^k \beta_r v_r + \Lambda w\right] \\ &= \sum_{r=1}^k \beta_r v_r + \frac{1}{N_i}\sum_{j=0}^{N_i-1} M^j(M-I)w \\ &= \sum_{r=1}^k \beta_r v_r + \frac{1}{N_i}\left[M^{N_i}w - w\right] \end{aligned}$$

Now since $\lim_{i\to\infty}\frac{1}{N_i}\left[M^{N_i}w - w\right] = 0$, it follows that $\lim_{i\to\infty} M_{N_i}v = \sum_{r=1}^k \beta_r v_r = \pi_{\text{ker}}v$. That is, $\lim_{i\to\infty} M_{N_i} = \pi_{\text{ker}}$.

Since every such convergent subsequence has the *same* limit, π_{ker}, we know that M_N converges. It must necessarily converge to the same limit as any of its subsequences. That is, it must also converge to π_{ker}.

Finally, we argue that M^∞ is Markov. Observe that M_N is Markov for all N. It is non-negative by definition, and $JM_N = J\frac{1}{N}\sum_{j=0}^{N-1} M^j = \frac{1}{N}\sum_{j=0}^{N-1} JM^j = \frac{1}{N}\sum_{j=0}^{N-1} J = J$. Now since $M_N \geq 0$, $M^\infty = \lim_{N\to\infty} M_N \geq \lim_{N\to\infty} 0 = 0$. Moreover, $JM^\infty = J\lim_{N\to\infty} M_N = \lim_{N\to\infty} JM_N = \lim_{N\to\infty} J = J$. □

We close this chapter with our first structure theorem in which we characterize the stable distributions of M in terms of M^∞. Specifically, we show that each column of M^∞

is a stable distribution of M. This result immediately implies that every Markov matrix, M, has a stable distribution.

Corollary 3.4. *For any Markov matrix, M, $\mathrm{stab}(M) = \mathrm{stab}\,(M^\infty) = M^\infty\Delta_n$. In particular, for all i, $M^\infty e_i \in M^\infty\Delta_n = \mathrm{stab}\,(M^\infty) = \mathrm{stab}(M)$.*

Proof. By Theorem 3.3 and Lemma 3.2, $M^\infty = \pi_{\mathrm{ker}}$ and $I - M^\infty = \pi_{\mathrm{im}}$, so that $\ker\Lambda^\infty = \ker\left(I - M^\infty\right) = \ker\pi_{\mathrm{im}}$. Then, by Lemma 2.10 iii), $\ker\Lambda^\infty = \ker\Lambda$. Therefore, $\mathrm{stab}\,(M^\infty) = \ker\Lambda^\infty \cap \Delta_n = \ker\Lambda \cap \Delta_n = \mathrm{stab}(M)$.

By Theorem 3.3, M^∞ is Markov, so that $M^\infty\Delta_n \subset \Delta_n$. Since $M^\infty\Delta_n \subset \mathrm{im}\,M^\infty = \mathrm{im}\,\pi_{\mathrm{ker}} = \ker\Lambda$, it follows that $M^\infty\Delta_n \subset \ker\Lambda \cap \Delta_n = \mathrm{stab}(M)$. Conversely, if $v \in \mathrm{stab}(M)$, since $v \in \ker\Lambda$, $M^\infty v = \pi_{\mathrm{ker}} v = v$, by Lemma 3.1 c) ii). Thus, $v \in M^\infty\Delta_n$, since $v \in \Delta_n$. □

Example 3.5. For example, consider $M = \begin{pmatrix} 0 & 1 & 0 & \frac{1}{6} \\ 1 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$. By induction, one may easily check that

$$M^i = \begin{pmatrix} [i \text{ is even}] & [i \text{ is odd}] & 0 & \frac{1}{3}\left(1 - \frac{1}{2^i}\right) \\ [i \text{ is odd}] & [i \text{ is even}] & 0 & \frac{1}{3}\left(1 - \frac{1}{2^i}\right) \\ 0 & 0 & 1 & \frac{1}{3}\left(1 - \frac{1}{2^i}\right) \\ 0 & 0 & 0 & \frac{1}{2^i} \end{pmatrix} \text{ and}$$

$$\frac{1}{N}\sum_{i=0}^{N-1} M^i = \begin{pmatrix} \frac{1}{2} + \frac{[N \text{ is odd}]}{2N} & \frac{1}{2} - \frac{[N \text{ is odd}]}{2N} & 0 & \frac{1}{3} + \frac{1-2^N}{3N2^{N-1}} \\ \frac{1}{2} - \frac{[N \text{ is odd}]}{2N} & \frac{1}{2} + \frac{[N \text{ is odd}]}{2N} & 0 & \frac{1}{3} + \frac{1-2^N}{3N2^{N-1}} \\ 0 & 0 & 1 & \frac{1}{3} + \frac{1-2^N}{3N2^{N-1}} \\ 0 & 0 & 0 & \frac{2^N-1}{N2^{N-1}} \end{pmatrix}$$

so that $M^\infty = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{pmatrix}$. In particular, we see that $M^\infty e_4 = \frac{2}{3}M^\infty e_1 + \frac{1}{3}M^\infty e_3$.

Since $M^\infty e_1 = M^\infty e_2 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix}$ and $M^\infty e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ are clearly stable distributions of M, all the columns of M^∞ are in $\mathrm{stab}(M)$, as Corollary 3.4 predicts. □

Chapter 4

Sub-Markov Matrix Invertibility

We begin this chapter by constructing a sets of walks in a directed, weighted graph corresponding to a given matrix. We then show how by aggregating the weights of such walks we arrive at an alternate formula for the powers of a matrix. Finally, we apply these tools to Markov matrices to obtain results about the invertibility of sub-Markov matrices.

4.1 Sequences and Walks

Let $\mathcal{S}_n(l) = \left\{\sigma : S_l^0 \to S_n\right\}$ be the set of sequences in S_n of length $l+1$. Likewise, let $\mathcal{S}_n(i,j,l) = \{\sigma \in \mathcal{S}_n(l) \mid \sigma_0 = j \text{ and } \sigma_l = i\}$ be the set of sequences in S_n of length $l+1$ starting with j and ending with i. For any set $s \in S_n$, we also define $\mathcal{S}_n(s,i,j,l) = \{\sigma \in \mathcal{S}_n(i,j,l) \mid \sigma_t \in s, \forall\, 0 < t < l\}$ to be the set of sequences in S_n of length $l+1$ from j to i that include elements of s only. Note that j and i do not themselves need to be in s, so for any s, $\mathcal{S}_n(s,i,j,1) = \mathcal{S}_n(i,j,1)$. Finally, $\mathcal{S}_n(s,i,j) = \bigcup_{l=1}^{\infty} \mathcal{S}_n(s,i,j,l)$ is the set of all such sequences of arbitrary length (greater than 1).

For any non-negative matrix, $M \geq 0$, and for any set $s \subset S_n$, we define $\mathcal{P}_M(s,i,j,l) \subset \mathcal{S}_n(s,i,j,l)$ as follows:

$$\mathcal{P}_M(s,i,j,l) = \left\{\sigma \in \mathcal{S}_n(s,i,j,l) \mid M_{\sigma_{t+1},\sigma_t} \neq 0,\, \forall 0 \leq t < l\right\}.$$

This represents the set of walks of length l in $G(M)$ from v_j to v_i that include vertices in V_s only. In addition, $\mathcal{P}_M(s,i,j) = \bigcup_{l=1}^{\infty} \mathcal{P}(s,i,j,l)$ is the set of all such walks of arbitrary length (greater than or equal to 1). There are corresponding definitions for $\mathcal{P}_M(i,j,l)$ and

$\mathcal{P}_M(l)$, representing the set of walks in $G(M)$ of length l that start at v_j and end at v_i, and the set of walks in $G(M)$ of length l, respectively.

Lemma 4.1. *For any $n \times n$ Markov matrix, M, for any set $s \subset S_n$ that is open in M, and for any $j \in s$, $\bigcup_{i \in s} \mathcal{P}_M(s,i,j,n)$ is a proper subset of $\bigcup_{i \in S_n} \mathcal{P}_M(i,j,n)$.*

Proof. Certainly $\bigcup_{i \in s} \mathcal{P}_M(s,i,j,n)$ is a subset of $\bigcup_{i \in S_n} \mathcal{P}_M(i,j,n)$, since $\bigcup_{i \in S_n} \mathcal{P}_M(i,j,n)$ corresponds to walks in $G(M)$ of length n starting at v_j, and $\bigcup_{i \in s} \mathcal{P}_M(s,i,j,n)$ corresponds to those same walks with the additional condition that they end in s and include only vertices in s.

Since V_s is open, by Lemma 1.2, from any vertex in V_s there is a walk σ that starts at v_j and ends at some vertex $v_k \notin V_s$. Since M is Markov, every vertex in $G(M)$ must have at least one outgoing edge, so σ can be extended to a walk σ' of length n. But then $\sigma' \in \bigcup_{i \in S_n} \mathcal{P}_M(i,j,n)$, and $\sigma' \notin \bigcup_{i \in s} \mathcal{P}_M(s,i,j,n)$, since it contains $v_k \notin V_s$. Thus $\bigcup_{i \in s} \mathcal{P}_M(s,i,j,n)$ is a proper subset of $\bigcup_{i \in S_n} \mathcal{P}_M(i,j,n)$. □

4.2 Matrix Powers and Walks

Given an $n \times n$ matrix, M, and a sequence of indices, $\sigma \in \mathcal{S}_n(l)$, $l \geq 1$ we will define

$$W(M,\sigma) \equiv \prod_{k=1}^{l} M_{\sigma_k,\sigma_{k-1}} \,. \tag{4.1}$$

The matrix entry $M_{\sigma_k,\sigma_{k-1}}$ corresponds to the weight on the edge from $v_{\sigma_{k-1}}$ to v_{σ_k} in $G(M)$. So the function $W(M,\sigma)$ has a graph-theoretic interpretation as the "total" weight of the walk σ in $G(M)$, where we aggregate weights by multiplication. Sequences of indices of length 1 correspond to walks of length 0, so following the usual convention that a product over an empty set is 1, $W(M,\sigma) = 1$ whenever $\sigma \in \mathcal{S}_n(0)$.

Lemma 4.2. *For any $n \times n$ matrix, M, and $\sigma \in \mathcal{S}_n(l)$, $W(M,\sigma) \neq 0$ iff σ is a walk in the graph $G(M)$, that is, iff $\sigma \in \mathcal{P}_M(l)$. Additionally, for any $i,j \in S_n$,*

$$\sum_{\sigma \in \mathcal{S}_n(i,j,l)} W(M,\sigma) = \sum_{\sigma \in \mathcal{P}_M(i,j,l)} W(M,\sigma).$$

Proof. If $\sigma \in \mathcal{P}_M(l)$, then $M_{\sigma_i,\sigma_{i-1}} \neq 0$, for all $1 \leq i \leq l$. Since $W(M,\sigma)$ is the product of non-zero values, it itself is not zero. Conversely, if $W(M,\sigma) \neq 0$, all of the terms in the product must be non-zero, so σ must be a walk of length l, i.e., $\sigma \in \mathcal{P}_M(l)$.

By definition, $\mathcal{P}_M(i,j,l) \subset \mathcal{S}_n(i,j,l) \subset \mathcal{S}_n(l)$. By the above argument, $W(M,\sigma) = 0$, for $\sigma \in \mathcal{S}_n(i,j,l) \setminus \mathcal{P}_M(i,j,l)$. Thus, when summing over all $\sigma \in \mathcal{S}_n(i,j,l)$, we can drop the zero terms, so that $\sum_{\sigma\in\mathcal{S}_n(i,j,l)} W(M,\sigma) = \sum_{\sigma\in\mathcal{P}_M(i,j,l)} W(M,\sigma)$. □

We can now give a graph-theoretic interpretation of matrix powers in terms of walks in $G(M)$. In words, for any $n \times n$ matrix M, the $(i,j)^{\text{th}}$ entry in M^l can be computed as the sum of the "total" weights of all walks in $G(M)$ from j to i of length l.

Lemma 4.3. *For any $n \times n$ matrix M, $\left(M^l\right)_{i,j} = \sum_{\sigma\in\mathcal{P}_M(i,j,l)} W(M,\sigma)$.*

Proof. Using the facts that the standard basis vectors e_k can be used to isolate matrix entries ($M_{i,j} = e_i^t M e_j$) and decompose the identity matrix ($I = \sum_{k=1}^n e_k e_k^t$), we have

$$\begin{aligned}
\left(M^l\right)_{i,j} &= e_i^t M^l e_j \\
&= e_i^t \underbrace{M \dots M}_{l \text{ times}} e_j \\
&= e_i^t M I M \dots M I M e_j \\
&= e_i^t M \left(\sum_{r_{l-1}=1}^n e_{r_{l-1}} e_{r_{l-1}}^t\right) M \dots M \left(\sum_{r_1=1}^n e_{r_1} e_{r_1}^t\right) M e_j \\
&= \sum_{r_{l-1}=1}^n \dots \sum_{r_1=1}^n e_i^t M e_{r_{l-1}} e_{r_{l-1}}^t M \dots M e_{r_1} e_{r_1}^t M e_j \\
&= \sum_{r_{l-1}=1}^n \dots \sum_{r_1=1}^n M_{i,r_{l-1}} M_{r_{l-1},r_{l-2}} \dots M_{r_2,r_1} M_{r_1,j} \,.
\end{aligned}$$

We now apply the substitution $\sigma(k) = r_k$, $\sigma(0) = j$, $\sigma(l) = i$, so that each choice of values for the summation variables, $\{r_1, \dots, r_{l-1}\}$, represents a unique choice of $\sigma : S_l^0 \to S_n$, a sequence of length $l+1$ which starts at j and ends at i. So this set of σs is precisely $\mathcal{S}_n(i,j,l)$. Thus,

$$\begin{aligned}
\left(M^l\right)_{i,j} &= \sum_{\sigma\in\mathcal{S}_n(i,j,l)} M_{\sigma_l,\sigma_{l-1}} M_{\sigma_{l-1},\sigma_{l-2}} \dots M_{\sigma_2,\sigma_1} M_{\sigma_1,\sigma_0} \\
&= \sum_{\sigma\in\mathcal{S}_n(i,j,l)} W(M,\sigma) \\
&= \sum_{\sigma\in\mathcal{P}_M(i,j,l)} W(M,\sigma) \,.
\end{aligned}$$

The final equality follows from Lemma 4.2. □

4.3 Sub-Markov Matrix Invertibility

We will now show that the laplacian, $\Lambda_{s,s}$, of a sub-Markov matrix, $M_{s,s}$, is invertible iff s is open with respect to M. The first half of this theorem will be crucial for specifying the reduction presented in Chapter 5, and its generalizations presented in Chapter 7. The second half will come into play when we make the connection between the reduction on Markov matrices and the associated construction on Markov chains, in section 5.3.

First, we must prove the following technical lemma relating walks in the graph of a principal submatrix, $M_{s,s}$, and certain walks in the graph of M.

Lemma 4.4. *For any $n \times n$ Markov matrix, M, for any set $s \subset S_n$,*

$$\sum_{\sigma \in \mathcal{P}_{M_{s,s}}(i,j,l)} W(M_{s,s}, \sigma) = \sum_{\sigma' \in \mathcal{P}_M(s,s_i,s_j,l)} W(M, \sigma')$$

Proof. Recall that because $M_{s,s}$ contains only the rows and columns of M with indices in s, $(M_{s,s})_{i,j} = M_{s_i,s_j}$. For any $\sigma \in \mathcal{P}_{M_{s,s}}(i,j,l)$, $\sigma_0 = j$ and $\sigma_l = i$. So $W(M_{s,s}, \sigma) = \prod_{k=1}^{l} (M_{s,s})_{\sigma_k,\sigma_{k-1}} = (M_{s,s})_{i,\sigma_{l-1}} \dots (M_{s,s})_{\sigma_1,j} = M_{s_i,s_{\sigma_{l-1}}} \dots M_{s_{\sigma_1},s_j}$. Now we can define a new σ' such that $\sigma'_0 = s_j$, $\sigma'_l = s_i$, and $\sigma'_k = s_{\sigma_k}$ for all $1 \leq k \leq l-1$. Since all interior indices σ' are in s, $\sigma' \in \mathcal{P}_M(s, s_i, s_j, l)$. Moreover, $W(M_{s,s}, \sigma) = W(M, \sigma')$.

We can also go the other direction. For any $\sigma' \in \mathcal{P}_M(s, s_i, s_j, l)$, there is a corresponding $\sigma \in \mathcal{P}_{M_{s,s}}(i,j,l)$ given by $\sigma_k = s^{-1}(\sigma'_k)$ for $0 \leq k \leq l$. Similarly, $W(M, \sigma') = W(M_{s,s}, \sigma)$. So there is a one-to-one correspondence between elements $\sigma \in \mathcal{P}_{M_{s,s}}(i,j,l)$ and elements $\sigma' \in \mathcal{P}_M(s, s_i, s_j, l)$ with $W(M_{s,s}, \sigma) = W(M, \sigma')$, and, therefore,

$$\sum_{\sigma \in \mathcal{P}_{M_{s,s}}(i,j,l)} W(M_{s,s}, \sigma) = \sum_{\sigma' \in \mathcal{P}_M(s,s_i,s_j,l)} W(M, \sigma')$$

□

Theorem 4.5. *If M is an $n \times n$ Markov matrix, with a principal submatrix, $M_{s,s}$, defined by an open set of indices, $s \subset S_n$, then*

a) For all integers $i \geq 0$, $\left\| M^i_{s,s} \right\|_1 \leq c^{\lfloor \frac{i}{n} \rfloor}$ for some $0 \leq c < 1$, and $\lim_{i \to \infty} M^i_{s,s} = 0$.

b) $I - M_{s,s}$ is invertible, and $\left(I - M_{s,s}\right)^{-1} = \sum_{i=0}^{\infty} M^i_{s,s}$.

Inversely, if s contains an entire closed class of M, then

c) $I - M_{s,s}$ *is not invertible, and*

d) $\lim_{i\to\infty} M_{s,s}^i \neq 0$.

Proof.

Proof of part a): Let $c = \left\|M_{s,s}^n\right\|_1 \geq 0$. We first show that $\left\|M_{s,s}^i\right\|_1$ is bounded above by $c^{\lfloor \frac{i}{n} \rfloor}$. Since M is Markov, and since $M_{s,s}$ is a submatrix of M, $0 \leq \left\|M_{s,s}\right\|_1 \leq 1$. Since the matrix norm is sub-multiplicative, for all i, $\left\|M_{s,s}^{i+1}\right\|_1 \leq \left\|M_{s,s}^i\right\|_1 \left\|M_{s,s}\right\|_1 \leq \left\|M_{s,s}^i\right\|_1$, so the sequence $\left\|M_{s,s}^i\right\|_1$, $i \geq 1$, is decreasing. Looking at every n^{th} term, we have the subsequence $\left\|M_{s,s}^{nk}\right\|_1 \leq \left(\left\|M_{s,s}^n\right\|_1\right)^k = c^k$, $k \geq 0$. Setting $k = \lfloor \frac{i}{n} \rfloor$, $nk \leq i$, so $\left\|M_{s,s}^i\right\|_1 \leq \left\|M_{s,s}^{nk}\right\|_1 \leq c^k = c^{\lfloor \frac{i}{n} \rfloor}$.

To prove that $c < 1$, we will show that the j^{th} column sum of $M_{s,s}^n$ (i.e., the s_j^{th} column sum of M) is strictly less than 1, so that all column sums of $M_{s,s}^n$ are strictly less than 1:

$$
\begin{aligned}
1 &= \sum_{i \in S_n} \left(M_{i,s_j}\right)^n && (4.2)\\
&= \sum_{i \in S_n} \sum_{\sigma \in \mathcal{P}_M(i,s_j,n)} W(M,\sigma) && (4.3)\\
&> \sum_{i \in s} \sum_{\sigma \in \mathcal{P}_M(s,i,s_j,n)} W(M,\sigma) && (4.4)\\
&= \sum_{i \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_M(s,s_i,s_j,n)} W(M,\sigma)\\
&= \sum_{i \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_{M_{s,s}}(i,j,n)} W(M_{s,s},\sigma)\\
&= \sum_{i \in S_{|s|}} \left(M_{s,s}^n\right)_{i,j} . && (4.5)
\end{aligned}
$$

Equation 4.2 follows from the fact that $JM = J$, so that $JM^n = J$, i.e., all column sums of M^n equal 1. Equation 4.3 follows from Lemma 4.3. Equation 4.4 follows from Lemma 4.1. Since $\bigcup_{i\in s} \mathcal{P}_M(s,i,s_j,n)$ is a proper subset of $\bigcup_{i\in S_n} \mathcal{P}_M(i,s_j,n)$, when we restrict the sum we throw away some positive terms. Finally, we rewrite the equation in terms of walks in $G\left(M_{s,s}\right)$ to arrive at Equation 4.5. Therefore, the j^{th} column sum of $M_{s,s}^n < 1$ for arbitrary j, so $\left\|M_{s,s}^n\right\|_1 < 1$.

Now, since $0 \leq c = \left\|M_{s,s}^n\right\|_1 < 1$, it follows that $\lim_{i\to\infty} c^{\lfloor \frac{i}{n} \rfloor} = 0$. Further, since

$0 \leq \left\|M_{s,s}^i\right\|_1 \leq c^{\lfloor \frac{i}{n} \rfloor}$, it follows that $\lim_{i\to\infty} \left\|M_{s,s}^i\right\|_1 = 0$. Finally, since the entries of $M_{s,s}^i$ are non-negative and bounded above by $\left\|M_{s,s}^i\right\|_1$, $\lim_{i\to\infty} M_{s,s}^i = 0$.

Proof of part b): By part a), $\sum_{i=0}^{\infty} \left\|M_{s,s}^i\right\|_1 \leq \sum_{i=0}^{\infty} c^{\lfloor \frac{i}{n} \rfloor}$. If we set $k = \lfloor \frac{i}{n} \rfloor$, then every n consecutive terms in this sum (for any $p \in \mathbb{N}$, $pn \leq i < (p+1)n$) can be grouped by a constant k (specifically, $k = p$), so that $\sum_{i=0}^{\infty} c^{\lfloor \frac{i}{n} \rfloor} = n \sum_{k=0}^{\infty} c^k = \frac{n}{1-c}$. Hence, the summation $\sum_{i=0}^{\infty} \left\|M_{s,s}^i\right\|_1$ is bounded above by a convergent series, and so it converges. Further, since the entries of $M_{s,s}^i$ are all non-negative and bounded above by $\left\|M_{s,s}^i\right\|_1$, the summation $\sum_{i=0}^{\infty} M_{s,s}^i$ also converges.

Next, we argue that $\sum_{i=0}^{\infty} M_{s,s}^i$ is the inverse of $\big(I - M_{s,s}\big)$. Since both matrices are square, by Theorem A.2, it suffices to show that the sum is a right-inverse, as follows:

$$\begin{aligned}
\big(I - M_{s,s}\big) \sum_{i=0}^{\infty} M_{s,s}^i &= \big(I - M_{s,s}\big) \lim_{j\to\infty} \sum_{i=0}^{j-1} M_{s,s}^i \\
&= \lim_{j\to\infty} \sum_{i=0}^{j-1} (I - M_{s,s}) M_{s,s}^i \\
&= \lim_{j\to\infty} \sum_{i=0}^{j-1} (M_{s,s}^i - M_{s,s}^{i+1}) \\
&= \lim_{j\to\infty} (M_{s,s}^0 - M_{s,s}^j) \\
&= I - \lim_{j\to\infty} M_{s,s}^j \\
&= I\,.
\end{aligned}$$

The final step in this derivation follows by part a).

Proof of part c): Inversely, assume s contains an entire closed class of M. By Theorem 2.11 a), $\ker \Lambda_{s,s} \neq 0$. But $(I - M_{s,s}) = (-1)(M_{s,s} - I) = (-1)\Lambda_{s,s}$, so $\ker(I - M_{s,s}) = \ker \Lambda_{s,s} \neq 0$, and $I - M_{s,s}$ is not invertible.

Proof of part d): As above, by Theorem 2.11 a), $\ker \Lambda_{s,s} \neq 0$. So there exists $v \in \ker \Lambda_{s,s}$ with $v \neq 0$. Now $M_{s,s} v = v$, so $(\lim_{i\to\infty} M_{s,s}^i)v = \lim_{i\to\infty}(M_{s,s}^i v) = \lim_{i\to\infty} v = v$. But $v \neq 0$. Therefore, $\lim_{i\to\infty} M_{s,s}^i \neq 0$.

□

Chapter 5

Two Useful Constructions

In this chapter, we present our two fundamental constructions, scaling and reduction. We will show that the result of these constructions applied to a Markov matrix is another Markov matrix which is, in a certain sense, equivalent, in that the stable distributions of the former can be recovered from those of the latter. Appropriate generalizations of these constructions to perturbed Markov matrices will form the basis of our main algorithm, presented in Chapter 7.

In addition, the reduction construction will allow us to sharpen the structure theorem shown in Corollary 3.4 by proving Theorem 5.17 and Corollary 5.18. We will also give a Markov chain interpretation of the construction in section 5.3. This will allow us to prove that the construction "composes" well; that is, if we use it to first eliminate one set of states, s_1, and then proceed to eliminate an additional set of states, s_2, we could obtain the same result by simply eliminating all the states, $s_1 \cup s_2$, at once.

5.1 Scaling

In this section, we introduce the notion of D-equivalent Markov matrices, where D is a matrix satisfying certain conditions. We will see that, intuitively, if we are only interested in stable distributions, we may replace any Markov matrix with one that is D-equivalent.

Definition 5.1. *If M_1 and M_2 are Markov matrices, we will say that M_2 is* D-equivalent *to M_1, and write $M_2 \approx_D M_1$, iff $D \geq 0$ has a non-negative left-inverse and $\ker \Lambda_1 = D \ker \Lambda_2$. When $D = I$, we will say that M_2 is* equivalent *to M_1, and write $M_2 \approx M_1$.*

In the next Lemma we see that that that for two matrices which are D-equivalent, D induces a 1-1 correspondence (i.e., a bijective mapping) between their sets of stable distributions. We will justify the term "equivalent" in Section 5.2, by showing that the other direction holds; that is, two matrices are equivalent iff they have the same set of stable distributions.

Lemma 5.2. *If M_1 and M_2 are Markov matrices, $M_2 \approx_D M_1$, with $D, L \geq 0$ and $LD = I$, then $D^*(v) = \frac{Dv}{\|Dv\|_1}$ is a bijective mapping from $\mathrm{stab}\,(M_2)$ to $\mathrm{stab}\,(M_1)$.*

Proof. First, observe that D^* maps $\mathrm{stab}\,(M_2)$ to $\mathrm{stab}\,(M_1)$. The matrix D maps $\ker M_2$ to $\ker M_1$, while the mapping D^* normalizes that result, dividing by $\|Dv\|_1$. Since D is non-negative, the image of $\mathrm{stab}\,(M_2)$ under D^* is non-negative, norm-1 vectors in $\ker M_1$. Next, we will show that D^* is bijective.

Injective: If $D^*(v) = D^*(w)$ for $v, w \in \mathrm{stab}\,(M_2)$, then $Dv = kDw$ for $k = \frac{\|Dv\|_1}{\|Dw\|_1} > 0$. So, $0 = Dv - kDw = D(v - kw)$. Further, $0 = LD(v - kw) = v - kw$. Hence, $v = kw$. But v and w are distributions, so $1 = Jv = Jkw = kJw = k$, and $v = w$. Thus, D^* is injective from $\mathrm{stab}\,(M_2)$ into $\mathrm{stab}\,(M_1)$.

Surjective: For any $w \in \mathrm{stab}\,(M_1) \subset \ker(\Lambda_1)$, since M_2 is D-equivalent to M_1, $w = Du$ for some $u \in \ker(\Lambda_2)$. Let $v = \frac{u}{\|u\|_1}$. $u = LDu = Lw \geq 0$, so $v \in \mathrm{stab}\,(M_2)$. Now $Dv = \frac{Du}{\|u\|_1} = \frac{w}{\|u\|_1}$, and $D^*v = \frac{Dv}{\|Dv\|_1} = \frac{w/\|u\|_1}{\|w/\|u\|_1\|_1} = \frac{w/\|u\|_1}{\|w\|_1/\|u\|_1} = \frac{w}{\|w\|_1} = w$. Thus, D^* is surjective from $\mathrm{stab}\,(M_2)$ onto $\mathrm{stab}\,(M_1)$, and hence a bijection between $\mathrm{stab}\,(M_2)$ and $\mathrm{stab}\,(M_1)$. □

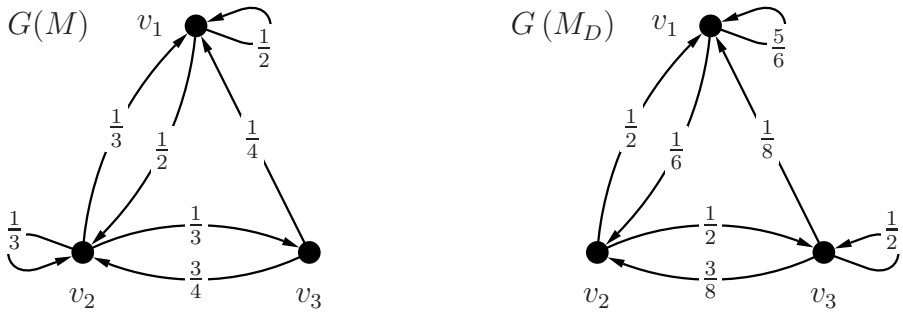
We now give a simple construction that operates on certain Markov matrices and produces D-equivalent results. For any Markov matrix, M, and any diagonal matrix, D, with $0 < D_{i,i}$ and $D_{i,i}\left(1 - M_{i,i}\right) \leq 1$ (i.e., $D_{i,i}$ is positive and sufficiently small) for all $i \in S_n$, we define $M_D = \Lambda D + I$ (and the corresponding $\Lambda_D = \Lambda D$). We say that M_D is the result of *scaling* M by D. Note that D is diagonal with positive diagonal entries; hence, it is invertible with positive inverse (it is easy to check that D^{-1} is also diagonal, with $D^{-1}_{i,i} = (D_{i,i})^{-1}$). In particular, it has both left and right inverses, and is thus both injective and surjective.

For example, if

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{3}{4} \\ 0 & \frac{1}{3} & 0 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}, \quad \text{then} \quad M_D = \begin{pmatrix} \frac{5}{6} & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{6} & 0 & \frac{3}{8} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

These correspond to the graphs in Table 5.1.

Table 5.1: $G(M)$ and $G\left(M_D\right)$ with $D = \text{diag}\left(\frac{1}{3}, \frac{3}{2}, \frac{1}{2}\right)$.



By looking at the corresponding graphs, we see that scaling by D adjusts the weight of the self-loop at each vertex with a proportional adjustment of the weights on the corresponding outgoing edges. When $D_{i,i}$ is close to 0, the weight of the self-loop at i is large (near 1), and when $D_{i,i}$ is close to $\left(1 - M_{i,i}\right)^{-1}$, the the weight of the self-loop is small (near 0). In Section 5.3, we will see that this may be viewed intuitively as adjusting the "diameter" of each vertex, in that the corresponding scaled Markov process spends either more or less time at that vertex.

Lemma 5.3. *Given a Markov matrix M, and any diagonal matrix, D, with $0 < D_{i,i}$ and $D_{i,i}\left(1 - M_{i,i}\right) \leq 1$ for all $i \in S_n$, M_D is a Markov matrix and $M_D \approx_D M$.*

Proof. We first show that M_D is Markov. Since $J\Lambda = 0$, $JM_D = J\Lambda D + J = J$, so the columns of M_D sum to 1. Moreover, all the off-diagonal entries of M_D are nonnegative, since, for $i \neq j$, $(M_D)_{ij} = (\Lambda D + I)_{i,j} = ((M - I)D)_{i,j} + I_{i,j} = (MD)_{i,j} - D_{i,j} = (MD)_{i,j}$ and $M, D \geq 0$, so $(MD)_{i,j} \geq 0$. Finally, we must show that all the diagonal entries of M_D are nonnegative. Observe that $(M_D)_{i,i} = (\Lambda D + I)_{i,i} = (\Lambda D)_{i,i} + I_{i,i} = \Lambda_{i,i}D_{i,i} + 1 = (M_{i,i} - 1)D_{i,i} + 1 = (-1)(1 - M_{i,i})D_{i,i} + 1$. Now $(1 - M_{i,i})D_{i,i} \leq 1$, so $(-1)(1 - M_{i,i})D_{i,i} \geq -1$, so $(M_D)_{i,i} \geq -1 + 1 = 0$.

Since D is non-negative with a non-negative inverse, D is surjective. So, by Lemma A.3 b), $\ker \Lambda = D \ker \Lambda D = D \ker \Lambda_D$. That is, $M_D \approx_D M$. □

We can use this scaling construction to produce infinitely many Markov matrices which are equivalent to a given Markov matrix. In particular, we have the following Corollary.

Corollary 5.4. *Given a Markov matrix M, if $0 < \epsilon$ and $\epsilon \max_i \left(1 - M_{i,i}\right) \leq 1$, then $M_\epsilon \equiv \Lambda\epsilon + I$ is a parameterized family of Markov matrices equivalent to M.*

Proof. Letting $D_\epsilon = \epsilon I$, $M_\epsilon = M_{D_\epsilon}$ so that Lemma 5.3 implies that $M_\epsilon \approx_{D_\epsilon} M$. But D is simply scalar multiplication. So by Lemma 5.3, $\ker \Lambda = D_\epsilon \ker \Lambda_\epsilon = \epsilon \ker \Lambda_\epsilon = \ker \Lambda_\epsilon$, and $M_\epsilon \approx M$. □

5.2 Reduction

In this section, we present a construction which allows us to "eliminate" an open set of indices, s, of a Markov matrix, M. Specifically, it produces a Markov matrix of strictly smaller dimension which is equivalent in the sense of Definition 5.1. In this way, it *reduces* the dimensionality of the matrix in a principled manner that does not lose any information regarding its long-term behavior. In Section 5.3, we will see that this corresponds directly to compressing the time spent in the corresponding states of a Markov chain to 0. Graphically, it effectively collapses the corresponding vertices in $G(M)$.

For M Markov with $s \in S_n$ an open set of indices in M, we may define:

$$\imath = P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} \end{pmatrix}, \tag{5.1}$$

$$p = \begin{pmatrix} I & -M_{\overline{s},s} \Lambda_{s,s}^{-1} \end{pmatrix} P_s^t, \text{ and} \tag{5.2}$$

$$\widehat{M} = p\Lambda\imath + I\,. \tag{5.3}$$

Example 5.5. In the Markov matrix $M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, the vertex with index 3 is an open set. We will take the reduction of M with respect to $s = \{3\}$.

s is already positioned in the lower-right corner of M, so the permutation matrices P_s and P_s^t are just the identity.

We have $\Lambda_{s,s} = \begin{pmatrix} -\frac{1}{2} \end{pmatrix}$, $\Lambda_{s,s}^{-1} = \begin{pmatrix} -2 \end{pmatrix}$, $M_{s,\overline{s}} = \begin{pmatrix} \frac{1}{4} \\ \\ \frac{1}{4} \end{pmatrix}$, $M_{\overline{s},s} = \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix}$.

So we can calculate $\imath = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$, $p = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}$,

$$\widehat{M} = p\Lambda\imath + I = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{4} \\ 1 & -1 & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{3}{4} \\ 1 & \frac{1}{4} \end{pmatrix}.$$

□

Theorem 5.6. *For M, $\imath$, p, and $\widehat{M}$ as defined above,*

a) $\imath, p \geq 0$ *and* $Jp = J$,

b) p *is surjective, and* $\imath$ *is injective.*

c) $\widehat{M} = M_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}}$,

d) $\widehat{M}$ *is Markov.*

Proof.

By Theorem 4.5 b),

$$-\Lambda_{s,s}^{-1} = \left(I - M_{s,s}\right)^{-1} = \lim_{i\to\infty} \sum_{j=0}^{i-1} M_{s,s}^{j} \geq 0$$

$M_{s,\overline{s}}$ and $M_{\overline{s},s}$ are also both non-negative, so p and $\imath$ are both well-defined and non-negative. We now show that the columns of p sum to 1. Since $P_s^t M P_s$ is Markov, $JP_s^t M P_s = J$. In particular, $JM_{s,s} + JM_{\overline{s},s} = J$. Therefore, $JM_{\overline{s},s} = J - JM_{s,s} = JI - JM_{s,s} = -J\Lambda_{s,s}$, so that

$$\begin{aligned} Jp &= J \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t = \begin{pmatrix} JI & J\Lambda_{s,s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t \\ &= \begin{pmatrix} JI & JI \end{pmatrix} P_s^t = JP_s^t = J \end{aligned}$$

Proof of part a): Proof of part b): Let $k = |s|$. Then p is $k \times n$ and $\imath$ is $n \times k$. Now p has rank k because its columns include the standard basis for $\mathbb{R}^k$. Similarly, $\imath$ has rank

k because the columns of $\imath^t$ include the standard basis for $\mathbb{R}^k$, and $\operatorname{rk}\imath = \operatorname{rk}\imath^t$. Then $\operatorname{rk} p + \dim\ker p^t = k$, so $\dim\ker p^t = 0$, and p is surjective. Similarly, $\operatorname{rk}\imath + \dim\ker\imath = k$, so $\dim\ker\imath = 0$, and $\imath$ is injective.

Proof of part c):

$$\begin{aligned}
\widehat{M} &= p\Lambda\imath + I \\
&= \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t \Lambda P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1}\,M_{s,\overline{s}} \end{pmatrix} + I \\
&= \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} \begin{pmatrix} \Lambda_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & \Lambda_{s,s} \end{pmatrix} \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1}\,M_{s,\overline{s}} \end{pmatrix} + I \\
&= \begin{pmatrix} \Lambda_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}} & 0 \end{pmatrix} \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1}\,M_{s,\overline{s}} \end{pmatrix} + I \\
&= \Lambda_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}} + I \\
&= M_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}}
\end{aligned}$$

Proof of part d): Since $-\Lambda_{s,s}^{-1}, M_{\overline{s},\overline{s}}, M_{\overline{s},s}, M_{s,\overline{s}} \geq 0$, $\widehat{M} \geq 0$. The columns of $\widehat{M}$ also sum to 1, since $J\widehat{M} = J(p\Lambda\imath + I) = J\Lambda\imath + J = J$, because $J\Lambda = 0$. So $\widehat{M}$ is Markov.

□

This motivates the following definition.

Definition 5.7. *The* reduction *of M with respect to s is the triple,* $(\widehat{M}, p, \imath)$.

We will refer to p and $\imath$ as the *projection* and *inclusion* operators of the reduction, since they are surjective and injective mappings, respectively. We will also sometimes refer to $\widehat{M}$ itself as the reduction.

Now we will examine the effect of this construction on the corresponding graphs. We will show that the entries of $\widehat{M}$ may be identified with walks of length at least 1 on $G_-(M)$ between vertices in $V_{\overline{s}}$ which only pass through vertices in V_s. [1]

For convenience, we will define $\mathcal{P}(M) \equiv (G_-(M))_T$, so that there is a path from u to v in $G_-(M)$ (or, equivalently, in $G(M)$) iff there is a walk from from u to v in $G_-(M)$ iff

[1] That is, whose interior/non-end vertices are in $V_{\overline{s}}$. In particular, this vacuously includes walks of length 1, since such walks have no interior vertices.

(u, v) is an edge in $\mathcal{P}(M)$. We will denote any of these equivalent propositions briefly by $(u, v) \in \mathcal{P}(M)$.

Theorem 5.8. *If M is Markov with $s \in S_n$ an open set of indices in M, $\left(\widehat{M}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j\right)} W(M,\sigma)$, and $\mathcal{P}\left(\widehat{M}\right) = \mathcal{P}(M)|_{V_{\overline{s}}}$, where v_i in $\mathcal{P}\left(\widehat{M}\right)_T$ corresponds to $v_{\overline{s}_i}$ in $\mathcal{P}(M)_T$.*

Proof. From Theorem 5.6 and Theorem 4.5 b), we have $\widehat{M} = M_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}} = M_{\overline{s},\overline{s}} + M_{\overline{s},s}\left(\sum_{l=0}^{\infty} M_{s,s}^l\right) M_{s,\overline{s}} = M_{\overline{s},\overline{s}} + \sum_{l=0}^{\infty} M_{\overline{s},s}\, M_{s,s}^l M_{s,\overline{s}}$. Therefore, $\left(\widehat{M}\right)_{i,j} = \left(M_{\overline{s},\overline{s}}\right)_{i,j} + \sum_{l=0}^{\infty}\left(M_{\overline{s},s}\, M_{s,s}^l M_{s,\overline{s}}\right)_{i,j}$.

By Lemma 4.3 and Lemma 4.4, $\left(M_{\overline{s},\overline{s}}\right)_{i,j} = \left(M^1_{\overline{s},\overline{s}}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_{M_{\overline{s},\overline{s}}}(i,j,1)} W(M_{\overline{s},\overline{s}},\sigma) = \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,1)} W(M,\sigma)$. We may require that the interior vertices of the walk σ lie in s ($\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,1)$ instead of $\mathcal{P}_M(\overline{s},\overline{s}_i,\overline{s}_j,1)$), since walks of length 1 have no interior vertices.

If we let $k = |s|$ and $\overline{k} = |\overline{s}|$, $M_{\overline{s},s}$ is $\overline{k} \times k$, $M_{s,s}^l$ is $k \times k$, and $M_{s,\overline{s}}$ is $k \times \overline{k}$. Applying matrix multiplication and Lemma 4.3,

$$\begin{aligned}
\left(M_{\overline{s},s}\, M_{s,s}^l M_{s,\overline{s}}\right)_{i,j} &= \sum_{q=1}^{k}\left((M_{\overline{s},s}M_{s,s}^l)_{i,q}(M_{s,\overline{s}})_{q,j}\right) \\
&= \sum_{q=1}^{k}\left(\sum_{p=1}^{k}\left((M_{\overline{s},s})_{i,p}(M_{s,s}^l)_{p,q}\right)(M_{s,\overline{s}})_{q,j}\right) \\
&= \sum_{q=1}^{k}\sum_{p=1}^{k}\left((M_{\overline{s},s})_{i,p}(M_{s,s}^l)_{p,q}(M_{s,\overline{s}})_{q,j}\right) \\
&= \sum_{q=1}^{k}\sum_{p=1}^{k}\left((M_{\overline{s},s})_{i,p}\left(\sum_{\sigma \in \mathcal{P}_{M_{s,s}}(p,q,l)} W(M_{s,s},\sigma)\right)(M_{s,\overline{s}})_{q,j}\right).
\end{aligned}$$

Looking at these terms in the scope of the entire matrix M,

$$\begin{aligned}
\left(M_{\overline{s},s}\, M_{s,s}^l M_{s,\overline{s}}\right)_{i,j} &= \sum_{q=1}^{k}\sum_{p=1}^{k}\left(M_{\overline{s}_i,s_p}\left(\sum_{\sigma \in \mathcal{P}_M(s,s_p,s_q,l)} W(M,\sigma)\right)M_{s_q,\overline{s}_j}\right) \\
&= \sum_{q=1}^{k}\sum_{p=1}^{k}\left(\sum_{\sigma \in \mathcal{P}_M(s,s_p,s_q,l)} M_{\overline{s}_i,s_p}M_{s_p,s_{\sigma_{l-1}}}\ldots M_{s_{\sigma_1},s_q}M_{s_q,\overline{s}_j}\right) \\
&= \sum_{\sigma' \in \mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,l+2)} W(M,\sigma').
\end{aligned}$$

By applying the substitution $\sigma'_0 = \overline{s}_j$, $\sigma'_1 = s_q$, $\sigma'_k = s_{\sigma_{s_{k-1}}}$ for $2 \leq k \leq l$, $\sigma'_{l+1} = s_p$, $\sigma'_{l+2} = \overline{s}_i$. Now σ' corresponds to a walk of length $l+2$, and all interior indices of σ' are in s, so $\sigma' \in \mathcal{P}_M(s, s_i, s_j, l+2)$, and since we are summing over all possibilities for p and q, the set of σ' is all of $\mathcal{P}_M(s, s_i, s_j, l+2)$.

Now, $\left(\widehat{M}\right)_{i,j} = \left(M_{\overline{s},\overline{s}}\right)_{i,j} + \sum_{l=0}^{\infty}\left(M_{\overline{s},s}\, M_{s,s}^{l} M_{s,\overline{s}}\right)_{i,j} = \sum_{\sigma\in\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,1)} W(M,\sigma) + \sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,l+2)} W(M,\sigma') = \sum_{l=1}^{\infty}\sum_{\sigma\in\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,l)} W(M,\sigma) = \sum_{\sigma\in\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j)} W(M,\sigma)$, since $\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j) = \bigcup_{l=1}^{\infty}\mathcal{P}_M(s,\overline{s}_i,\overline{s}_j,l)$.

If there is an edge, $(v_j, v_i) \in \mathcal{P}(M)$, for $i \neq j$, then there a walk of length at least 1 from v_j to v_i in $G(M)$. If $i, j \in \overline{s}$, this walk may then be decomposed into a concatenation of walks whose interior vertices are in V_s which originate and terminate in $V_{\overline{s}}$. Each of these walks correspond to an edge in $G\left(\widehat{M}\right)$, and together they make a walk in $G\left(\widehat{M}\right)$ and a single edge in $\mathcal{P}\left(\widehat{M}\right)$. Conversely, any edge between distinct vertices in $\mathcal{P}\left(\widehat{M}\right)$ corresponds to a walk in $G\left(\widehat{M}\right)$, which corresponds to a walk in $G(M)$ and an edge in $\mathcal{P}(M)$. Since, by definition, both graphs also contain all self-loops, $\mathcal{P}\left(\widehat{M}\right) = \mathcal{P}(M)|_{V_{\overline{s}}}$. □

Example 5.9. Recall that in Example 5.5, we calculated that the reduction of $M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ with respect to the open set $s = \{3\}$ is $\widehat{M} = \begin{pmatrix} 0 & \frac{3}{4} \\ 1 & \frac{1}{4} \end{pmatrix}$.

Now we can see that is the result we would expect from our graphical intuition. In M, there is no path with interior vertices in s from v_1 back to itself, so $\widehat{M}_{1,1} = 0$. There is one such path (of length one - it has no interior vertices) from v_1 to v_2 with weight 1, so $\widehat{M}_{2,1}$ = 1.

$\widehat{M}_{1,2}$ and $\widehat{M}_{2,2}$ are more complicated. There are infinitely many paths beginning at v_2 and ending at v_1 or v_2 with interior vertices in s, since there is a self-loop at v_3.

So $\widehat{M}_{1,2}$ is the sum of the weight of the edge $(2,1)$ and the weights of all paths which begin at 2, cycle at 3 i times, and end at 1. That is, $\widehat{M}_{1,2} = M_{1,2} + \sum_{i=0}^{\infty} M_{1,3}(M_{3,3})^i M_{3,2} = \frac{1}{2} + \sum_{i=0}^{\infty}\frac{1}{4}(\frac{1}{2})^i\frac{1}{2} = \frac{1}{2} + \frac{1}{4}\cdot 2\cdot\frac{1}{2} = \frac{3}{4}$.

Similarly, $\widehat{M}_{2,2} = \sum_{i=0}^{\infty} M_{2,3}(M_{3,3}^i)M_{3,2} = \sum_{i=0}^{\infty}\frac{1}{4}(\frac{1}{2})^i\frac{1}{2} = \frac{1}{4}$. □

Theorem 5.8 leads to the following important geometric property of the reduction construction. Intuitively, it says that the reduction of an open set is open.

Theorem 5.10. *If s and s' are open sets of indices of M such that $s \cup s'$ is also open, and $\widehat{M}$ is the reduction of M with respect to s, then $\overline{s}^{-1}\left(s'\right) \equiv \left\{j \in S_{|\overline{s}|} \mid \overline{s}_j \in s'\right\}$, the indices of $\widehat{M}$ that correspond to indices in s', is open with respect to $\widehat{M}$.*

Proof. Consider any if $j \in \overline{s}^{-1}\left(s'\right)$. By Lemma 1.2, since $s \cup s'$ is open with respect to M, and $\overline{s}_j \in s'$, there is an edge in $\mathcal{P}(M)$ from $v_{\overline{s}_j}$ to some $v_{\overline{s}_k}$, where $\overline{s}_k \in \overline{s \cup s'} = \overline{s} \cap \overline{s}'$. That is, by Theorem 5.8, there is an edge in $\mathcal{P}\left(\widehat{M}\right)$ from j to $k \notin \overline{s}^{-1}\left(s'\right)$. Hence, by Lemma 1.2, $\overline{s}^{-1}\left(s'\right)$ is open with respect to $\widehat{M}$. □

Corollary 5.11. *If M is unichain or irreducible, then for any open set, s, so is the corresponding reduction $\widehat{M}$.*

Proof. First, observe that M is unichain iff there is an open set, s, with $|s| = n - 1$. For example, if M is unichain and i is a chosen index in the closed class, then $s = S_n - \{i\}$ is open. Conversely, if there is an open set, s, with $|s| = n-1$, $\overline{s} = \{i\}$ can only be contained in at most one closed class. In particular, if M has more than one closed class, at least one of them must be contained in s. Since, by Lemma 1.1, M has at least one closed class, it must have exactly one, that is, M must be unichain.

Now assume that M is unichain and s is open. Then there must be some $i \in \overline{s}$ which is in its closed class. In particular, $S_n - \{i\}$ is open. If we take $s' \equiv S_n - \{i\} - s$, then we may apply Theorem 5.10 to conclude that $\overline{s}^{-1}\left(s'\right)$ is open with respect to $\widehat{M}$. Since $s' \subset \overline{s}$, $\left|s' \cap \overline{s}\right| = \left|s'\right| = |\overline{s}| - 1$, and $\widehat{M}$ is $|\overline{s}|$-dimensional, we may conclude that $\widehat{M}$ is unichain.

If M is irreducible, then $\mathcal{P}(M)$ must be complete. Since $\mathcal{P}\left(\widehat{M}\right) = \mathcal{P}(M)|_{V_{\overline{s}}}$, $\mathcal{P}\left(\widehat{M}\right)$ must be complete, as well. In particular, $\widehat{M}$ is irreducible. □

We will show in Section 5.3 that if we consider a Markov process, X_*, with transition matrix, M, and any initial distribution, $\widehat{M}$ corresponds to another Markov chain, $\widehat{X}_*$, which is just X_*, except that we pass through states of s without pause. We will likewise obtain a compelling probabilistic interpretation of p as a mapping from the initial distribution of X_* to that of $\widehat{X}_*$.

While there is no obvious probabilistic interpretation of $\imath$, it possesses several useful properties. Most importantly, the reduction construction "preserves" the kernel of the laplacian in the following sense.

Theorem 5.12. *Given a Markov matrix M and an open set of indices, s, using the notation of Theorem 5.6, $\imath$ has a non-negative left-inverse, $\pi_{\overline{s}}$, and $\ker \Lambda = \imath \ker \widehat{\Lambda}$, so that $\widehat{M} \approx_\imath M$.*

Proof. Consider the matrix $B = \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix}$. Multiply it on the left by three invertible matrices: $A_1 = \begin{pmatrix} I & M_{\overline{s},s} \\ 0 & I \end{pmatrix}$, which has inverse $\begin{pmatrix} I & -M_{\overline{s},s} \\ 0 & I \end{pmatrix}$, $A_2 = \begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s} \end{pmatrix}$, which has well-defined inverse $\begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s}^{-1} \end{pmatrix}$ because s is open (by Theorem 4.5, $\Lambda_{s,s} = -(I - M_{s,s})$ is invertible), and $A_3 = P_s$, which has inverse P_s^t.

$$\begin{aligned}
A_3A_2A_1B &= P_s \begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s} \end{pmatrix} \begin{pmatrix} I & M_{\overline{s},s} \\ 0 & I \end{pmatrix} \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} \\
&= P_s \begin{pmatrix} \widehat{\Lambda} & M_{\overline{s},s} \\ 0 & \Lambda_{s,s} \end{pmatrix} \\
&= P_s \begin{pmatrix} \Lambda_{\overline{s},\overline{s}} - M_{\overline{s},s}\Lambda_{s,s}^{-1}M_{s,\overline{s}} & M_{\overline{s},s} \\ 0 & \Lambda_{s,s} \end{pmatrix} \qquad (5.4) \\
&= P_s \begin{pmatrix} \Lambda_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & \Lambda_{s,s} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\overline{s}} & I \end{pmatrix} \\
&= P_s \left(P_s^t \Lambda P_s\right) \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\overline{s}} & I \end{pmatrix} \qquad (5.5) \\
&= \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\overline{s}} & I \end{pmatrix}.
\end{aligned}$$

Equation 5.4 follows from Theorem 5.6 c), and Equation 5.5 is rewritten according to the principal submatrix permutation of Λ. Now since A_1, A_2, A_3 are invertible, in particular they are injective. So by Lemma A.3 c), $\ker A_3A_2A_1B = \ker B$. That is, $\ker \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\overline{s}} & I \end{pmatrix} = \ker \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix}$.

Now consider Lemma 2.7 with $s' = S_n - S_{|\overline{s}|}$, so that $\imath_{\overline{s}'} = \begin{pmatrix} I \\ 0 \end{pmatrix}$, $\imath_{s'} = \begin{pmatrix} 0 \\ I \end{pmatrix}$, $\pi_{\overline{s}'} = \begin{pmatrix} I & 0 \end{pmatrix}$, and $\pi_{s'} = \begin{pmatrix} 0 & I \end{pmatrix}$. Since $\begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix} \widehat{\Lambda} \begin{pmatrix} I & 0 \end{pmatrix} +$

$\begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} 0 & I \end{pmatrix}$, we then have

$$\begin{aligned}
\ker \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} &= \ker \left[\begin{pmatrix} I \\ 0 \end{pmatrix} \widehat{\Lambda} \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} 0 & I \end{pmatrix} \right] \\
&= \ker \left[\imath_{\overline{s}'} \widehat{\Lambda} \pi_{\overline{s}'} + \imath_{s'} \pi_{s'} \right] \\
&= \ker \left[\imath_{\overline{s}'} \widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \ker \left[\imath_{s'} \pi_{s'} \right] && \text{by Lemma A.3 d)} \\
&= \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \ker \pi_{s'} && \text{by Lemma A.3 c)} \\
&= \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \operatorname{im} \imath_{\overline{s}'} && \text{by Lemma 2.7 c)} \\
&= \imath_{\overline{s}'} \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \imath_{\overline{s}'} \right] && \text{by Lemma A.3 a)} \\
&= \imath_{\overline{s}'} \ker \widehat{\Lambda} \,. && \text{by Lemma 2.7 a)}
\end{aligned}$$

To summarize, we have $\ker \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix} = \imath_{\overline{s}'} \ker \widehat{\Lambda}$. P_s has inverse P_s^t, and $\begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix}$ has inverse $\begin{pmatrix} I & 0 \\ \Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix}$, so in particular both matrices are surjective, and therefore by Lemma A.3 b),

$$\begin{aligned}
\ker \Lambda &= P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix} \ker \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix} \\
&= P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix} \imath_{\overline{s}'} \ker \widehat{\Lambda} \\
&= P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} & I \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \ker \widehat{\Lambda} \\
&= P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} \end{pmatrix} \ker \widehat{\Lambda} \\
&= \imath \ker \widehat{\Lambda} \,.
\end{aligned}$$

It is easy to check that $\pi_{\overline{s}} = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix} = \pi_{\overline{s}'} P_s^t \geq 0$ is a left-inverse of $\imath$:

$$\pi_{\overline{s}} \imath \;=\; \pi_{\overline{s}'} P_s^t P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} \end{pmatrix} = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} \end{pmatrix} = I \,.$$

Now since $\imath$ has a non-negative left-inverse, and $\ker \Lambda = \imath \ker \widehat{\Lambda}$, $M \approx_\imath \widehat{M}$ by definition. □

Theorem 5.12 and Lemma 5.2 then give the following important result.

Corollary 5.13. *Given a Markov matrix M if s is an open set of indices, with corresponding reduction $\widehat{M}$, then $\imath^*(v) = \frac{\imath v}{\|\imath v\|_1}$ is a bijective mapping from* $\operatorname{stab}\left(\widehat{M}\right)$ *to* $\operatorname{stab}(M)$.

This allows us to give simple and direct proofs of important structure theorems for Markov matrices. For example, we may prove the uniqueness of stable distributions in a very general setting, without restrictive assumptions of aperiodicity or ergodicity, etc.

Theorem 5.14. *Given a Markov matrix M with k closed classes,* $\dim \ker \Lambda = k$.

Proof. Take any maximal, open set of indices, s. Then $\overline{s}$ must have exactly one element from each distinct closed class of M, so that $|\overline{s}| = k$. Now consider the reduction, $\widehat{M}$, with respect to s. Since there are no walks between closed classes, by Theorem 5.8 $\left(\widehat{M}\right)_{i,j} = 0$ for $i \neq j$. In particular, $\widehat{M} = I$ and $\ker \widehat{\Lambda} = \mathbb{R}^k$. Therefore, by Corollary 5.13, $\dim \ker \Lambda = k$. □

Corollary 5.15. *Every unichain Markov matrix M with closed class, s, has a unique stable distribution v such that $v_i \neq 0 \iff i \in s$.*

Proof. By Corollary 3.4, $|\operatorname{stab} M| > 1$. More specifically, by Lemma 2.10 d), $\operatorname{stab} M$ contains $\imath_s \overline{v}$, where $\overline{v} \in \operatorname{stab} M_{s,s}$, the stable distribution of the principal submatrix, $M_{s,s}$. Since M is unichain, by Theorem 5.14, $\dim \ker \Lambda = 1$. Thus, if $v, w \in \operatorname{stab} M \subset \ker \Lambda$, we would have $v = kw$. However, since $1 = Jv = kJw = k$, $v = w$, so that $|\operatorname{stab} M| = 1$, namely, $\operatorname{stab} M = \{\imath_s \overline{v}\}$.

Letting $v = \imath_s \overline{v}$, consider $s' = \{i \mid v_i \neq 0\}$. We first show that $s' \subset s$ by considering the contrapositive. If $i \notin s$, then $i = \overline{s}_j$ for some j. Therefore, $v_i = e_i^t v = e_{\overline{s}_j}^t v = e_j^t \pi_{\overline{s}} v = e_j^t \pi_{\overline{s}} \imath_s \overline{v} = 0$, by Lemma 2.7 a).

Since $v \in \ker \Lambda$, $v = Mv$. For any $i \in s - s'$, $0 = v_i = e_i^t v = e_i^t v = e_i^t M v = e_i^t M \sum_j v_j e_j = \sum_{j \in s'} v_j e_i^t M e_j$, since $v_j = 0$ for all $j \in s - s'$. Since $v_j > 0$ and $e_i^t M e_j \geq 0$ for all $j \in s'$, we must have $e_i^t M e_j = 0$ for all $j \in s'$. That is, there are no edges from $j \in s'$ to $i \in s - s'$ in $G(M)$. In particular, there are no walks from $j \in s'$ to $i \in s - s'$. Since s is an SCC, this is impossible, unless $s - s' = \emptyset$, i.e., $s' = s$. In other words, $v_i \neq 0 \iff i \in s$. □

This immediately gives the following well-known result Horn and Johnson (1985).

Corollary 5.16. *Every irreducible Markov matrix M has a unique stable distribution, $v > 0$.*

This leads to the following structure theorem for Markov matrices (cf., the proof of Theorem 2.1 in (Karlin and Taylor, 1981, p. 4)). The reduction construction provides a conceptually satisfying, constructive proof.

Theorem 5.17. *Given a Markov matrix M, with closed classes, $s^j,\ j = 1, \ldots, k$, let $\overline{M}_j$ be the principal submatrix on s^j with unique, stable distribution, $\overline{v}^j$. Let $v^j = \imath_{s^j} \overline{v}^j$.*

- $\ker \Lambda = \operatorname{span} \left\{ v^1, \ldots, v^k \right\}$*;*
- $\ker (\Lambda) = \operatorname{span} (\operatorname{stab} (M))$*; and*
- *every $v \in \operatorname{stab} M$ is a convex combination of $v^j = \imath_{s^j} \overline{v}^j$, i.e., $v = \sum_{j=1}^k \alpha_j v^j$ for $0 \leq \alpha_j \leq 1$ with $\sum_{j=1}^k \alpha_j = 1$.*

Proof. We should first observe that, by Lemma 2.10, $\overline{M}_i$ is Markov. Moreover, since $G_- \left(\overline{M}_i \right)$ is strongly connected, $\overline{M}_i$ is irreducible. Therefore, by Corollary 5.16, $\overline{M}_i$ does have a unique, stable distribution, $\overline{v}^j > 0$.

Now define $D = \begin{pmatrix} v^1 & \cdots & v^k \end{pmatrix} \geq 0$ and $L = \begin{pmatrix} J\pi_{s^1} \\ \vdots \\ J\pi_{s^k} \end{pmatrix} \geq 0$. Notice that

$$\begin{aligned} LD &= \begin{pmatrix} J\pi_{s^1}v_1 & \cdots & J\pi_{s^1}v_k \\ \vdots & \ddots & \vdots \\ J\pi_{s^k}v_1 & \cdots & J\pi_{s^k}v_k \end{pmatrix} \\ &= \begin{pmatrix} J\pi_{s^1}\imath_{s^1}\overline{v}^1 & \cdots & J\pi_{s^1}\imath_{s^k}\overline{v}^k \\ \vdots & \ddots & \vdots \\ J\pi_{s^k}\imath_{s^1}\overline{v}^1 & \cdots & J\pi_{s^k}\imath_{s^k}\overline{v}^k \end{pmatrix} \\ &= \begin{pmatrix} J\overline{v}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J\overline{v}^k \end{pmatrix} = I \end{aligned}$$

By Lemma 2.10, D maps $\mathbb{R}^k$ into $\ker \Lambda$. By Theorem 5.14, maps $\mathbb{R}^k = \ker I$ onto $\ker \Lambda$, so that $\ker \Lambda = \operatorname{im} D = \operatorname{span}\left\{v^1, \ldots, v^k\right\}$. Since $\left\{v^1, \ldots, v^k\right\} \subset \operatorname{stab} M \subset \ker M$, we may infer that $\operatorname{span}\left\{v^1, \ldots, v^k\right\} \subset \operatorname{span}\operatorname{stab} M \subset \ker M$. In particular, $\operatorname{span}\operatorname{stab} M \subset \ker M$.

Since D is left-invertible, it is also injective. In other words, $I \approx_D M$. In fact, since $v_j \in \operatorname{stab} M$, $JD = \begin{pmatrix} Jv_1 & \cdots & Jv_k \end{pmatrix} = J$, so that for any $v \in \operatorname{stab} I = \Delta_k$, $\|Dv\|_1 = JDv = Jv = 1$, and $D^* = D$ on Δ_k. That is, by Lemma 5.2, D gives a bijection of Δ_k with $\operatorname{stab} M$. In particular, element in $\operatorname{stab} M$ is a convex combination of the v^j. □

We may now give a further characterization of $\imath$ when s is maximal.

Corollary 5.18. *Given a Markov matrix M and any maximal, open set of indices, s, with corresponding reduction, $(\widehat{M}, p, \imath)$,*

- *there is a 1-1 correspondence, α, between the v^j from Theorem 5.17 and $S_{\overline{s}}$, such that $\overline{s}_{\alpha(j)} \in s^j$, and*
- *$v^j \propto \imath e_{\alpha(j)}$, the $\alpha(j)^{th}$ column of $\imath$.*

Proof. First, remember from the proof of Theorem 5.14, if we let $k = |\overline{s}|$, for each $j \in S_k$, $\left|\overline{s} \cap s^j\right| = 1$. Thus, there is a well-defined mapping, $\alpha : S_k \to S_k$, such that $\overline{s} \cap s^j = \left\{\overline{s}_{\alpha(j)}\right\}$. Moreover, $\widehat{M} = I$, and $\imath$ maps $\ker \widehat{\Lambda} = \mathbb{R}^k$ onto $\ker \Lambda$, i.e., $\operatorname{im} \imath = \ker \Lambda$. In particular, for a given j, $v^j = \imath\beta$, for some β.

But from Theorem 5.12, we know that $\pi_{\overline{s}}$ is a left-inverse for $\imath$, so that $\beta = \pi_{\overline{s}} v^j = \pi_{\overline{s}} \imath_{s^j} \overline{v}^j$. The components of β are $e_r^t \beta = e_r^t \pi_{\overline{s}} \imath_{s^j} \overline{v}^j = e_{\overline{s}_r}^t \imath_{s^j} \overline{v}^j$, which is 0 unless $\overline{s}_r \in s^j$, that is, $r = \alpha(j)$. Thus, β has exactly one non-zero component and $v^j = \imath\beta = \sum_{r \in S_k} \imath e_r e_r^t \beta = \imath e_{\alpha(j)} e_{\alpha(j)}^t \beta$. In particular, $v^j \propto \imath e_{\alpha(j)}$. □

Theorem 5.19. *Under the assumptions and notation of Corollary 5.18, if $\beta = \alpha^{-1}$, then*

- $\imath(p\imath)^{-1} = \begin{pmatrix} v^{\beta(1)} & \cdots & v^{\beta(k)} \end{pmatrix}$ *and*
- $M^\infty = \imath(p\imath)^{-1} p$.

Proof. Let s be a maximal, open set of states with corresponding reduction, $(\widehat{M}, p, \imath)$. From the proof of Theorem 5.14, we know that $\widehat{M} = I$. We will begin by deriving two additional properties of such a maximal reduction.

Since $I = \widehat{M} = p\Lambda\imath + I$, we see that $p\Lambda\imath = 0$. In fact, we may show that $p\Lambda = 0$, that is, $p(M - I) = 0$ or $pM = p$. Using the notation of Theorem 5.6,

$$\begin{aligned} pM &= \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t P_s \begin{pmatrix} M_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & M_{s,s} \end{pmatrix} P_s^t = \begin{pmatrix} M_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}} & M_{\overline{s},s} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M \end{pmatrix} \\ &= \begin{pmatrix} \widehat{M} & M_{\overline{s},s} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1}\left(\Lambda_{s,s} + I\right) \end{pmatrix} P_s^t = \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t = p \end{aligned}$$

In addition, we may show that the product, $p\imath$, is an invertible diagonal matrix. Again, using the notation of Theorem 5.6, $p\imath = \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\overline{s}} \end{pmatrix} = I + M_{\overline{s},s}\,\overline{\Lambda}^{-2} M_{s,\overline{s}}$. Since I is diagonal, it suffices to show that $M_{\overline{s},s}\,\Lambda_{s,s}^{-2} M_{s,\overline{s}}$ is diagonal with non-negative entries.

As in the proof of Theorem 5.8,

$$M_{\overline{s},s}\,\Lambda_{s,s}^{-2} M_{s,\overline{s}} = M_{\overline{s},s} \left(\sum_{k=0}^{\infty} M_{s,s}^k\right) \left(\sum_{l=0}^{\infty} M_{s,s}^l\right) M_{s,\overline{s}} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} M_{\overline{s},s}\, M_{s,s}^{k+l} M_{s,\overline{s}}$$

To prove that this is diagonal, it suffices to show that, for any $l \geq 0$ and $i \neq j$, $e_i^t M_{\overline{s},s}\, M_{s,s}^l M_{s,\overline{s}} e_j = 0$. As before, this is equal to $\sum_{\sigma \in \mathcal{S}_n\left(s, \overline{s}_i, \overline{s}_j, l+1\right)} M_\sigma$. As in the proof of Theorem 5.14, when

$i \neq j$, there are no walks in $G(M)$ from $\overline{s}_j$ to $\overline{s}_i$, since $\overline{s}_j$ and $\overline{s}_i$ are in different closed classes. Thus, this sum is 0. By Theorem 5.6, $Jp\imath = J\imath$, so that the diagonal entries of $p\imath$ correspond to the column sums of $\imath$. In particular, since $\imath \geq 0$, they are non-negative.

We now prove the first part of the theorem to give a formula for $\imath$. By the above discussion $\imath(p\imath)^{-1}$ is the result of dividing each column by its corresponding sum, so that the resulting columns are all distributions. By Corollary5.18, $v^r \propto \imath e_{\alpha(r)}$, so that $v^{\beta(j)} \propto \imath e_j$ and the j^{th} column of $\imath(p\imath)^{-1}$ must be $v^{\beta(j)}$.

Finally, we will now show that $\imath(p\imath)^{-1}p = M^\infty$ by appealing to Theorem 3.3. That is, we will show that $\imath(p\imath)^{-1}p = \pi_{\text{ker}}$. First, observe that, by Theorem 5.17 and Corollary 5.18, $\operatorname{im}\imath = \ker\Lambda$. By Lemma 3.1 i), for any $v \in \mathbb{R}^n$, $v = \pi_{\text{ker}}v + \pi_{\text{im}}v$. Since $\operatorname{im}\pi_{\text{ker}} = \operatorname{im}\imath$, we may write $\pi_{\text{ker}}v = \imath\alpha$ for some α. Likewise, since $p\Lambda = 0$, and $\operatorname{im}\pi_{\text{im}} = \operatorname{im}\Lambda$, $p\pi_{\text{im}}v = 0$. Thus, $\imath(p\imath)^{-1}pv = \imath(p\imath)^{-1}p\pi_{\text{ker}}v = \imath(p\imath)^{-1}p\imath\alpha = \imath\alpha = \pi_{\text{ker}}v$. In particular, $\imath(p\imath)^{-1}pv = \pi_{\text{ker}}v$. □

Example 5.20. Returning to Example 3.5 with $M = \begin{pmatrix} 0 & 1 & 0 & \frac{1}{6} \\ 1 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$, we may verify the conclusions of Theorem 5.19. This matrix has two closed classes, $\{1,2\}$ and $\{3\}$. Reducing M with respect to the maximal open set, $s = \{2,4\}$ yields the 2-dimensional identity matrix with $\imath = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $p = \begin{pmatrix} 1 & 1 & 0 & \frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{pmatrix}$. Then, $p\imath \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $\imath(p\imath)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$. The two columns of this matrix are clearly the stable distributions associated with the corresponding closed classes. Moreover, multiplying this on the right

by p yields $\imath(p\imath)^{-1}p = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{pmatrix} = M^{\infty}$, as previously computed □

We can also justify our definition of equivalence.

Corollary 5.21. *For any two Markov matrices,* $M_1 \approx M_2 \iff \mathrm{stab}\,(M_1) = \mathrm{stab}\,(M_2)$. *That is, two Markov matrices are equivalent if and only if they have the same set of stable distributions.*

Proof. If M_1 and M_2 are equivalent, then $\ker(\Lambda_1) = \ker(\Lambda_2)$ by definition, so that $\mathrm{stab}\,(M_1) = \ker(\Lambda_1) \cap \Delta_n = \ker(\Lambda_2) \cap \Delta_n = \mathrm{stab}\,(M_2)$. Conversely, if $\mathrm{stab}\,(M_1) = \mathrm{stab}\,(M_2)$, by Theorem 5.17, $\ker(\Lambda_1) = \mathrm{span}\,(\mathrm{stab}\,(M_1)) = \mathrm{span}\,(\mathrm{stab}\,(M_2)) = \ker(\Lambda_2)$. In particular, M_1 and M_2 are equivalent. □

5.3 A Markov Chain Interpretation of the Constructions

In this section, we will review the basic definitions regarding finite-state, stationary, Markov chains, assuming the reader is familiar with basic probability and measure theory. Our goal is to show how the construction of section 5.2 corresponds to a corresponding construction on finite-state, stationary Markov chains. A *discrete-time stochastic process* (or *chain*) is a sequence, $X_* \equiv \{X_t\}_{t=0}^{\infty}$, of random variables i.e., real-valued measurable functions on some shared probability space, (Ω, μ). As is common, we will write $\Pr[\omega]$ for the probability of a measurable subset $\omega \subset \Omega$. Likewise, given a random variable, X, we will write $\Pr[X \in \beta]$ for $\Pr[X^{-1}(\beta)]$, assuming that $\beta \in \mathcal{B}$, the so-called Borel sets of $\mathbb{R}$.[2] In this way, we avoid explicit reference to Ω and μ. We will also write $\Pr[X = x]$ for $\Pr[X \in \{x\}]$. The *support*, supp_X, of a random variable, X, is the smallest Borel set, β, such that $\Pr[X \in \beta] = 1$. In this paper, we will restrict attention to those chains whose *state space*, $\mathcal{S} = \bigcup_i \mathrm{supp}_{X_i}$, is a finite set, and we establish the convention that $\overline{\beta} \equiv \mathcal{S} - \beta$.

[2] $\mathcal{B}$ is the smallest collection of subsets of $\mathbb{R}$ which contains all half-intervals (i.e., $[a, \infty)$) and is closed under countable unions/intersections and taking complements; in particular, it contains all countable subsets.

A chain, X_*, is *Markov* iff for all t and $s_0, \ldots, s_{t+1} \in \mathcal{S}$, such that $\Pr[X_t = s_t, \ldots, X_0 = s_0] \neq 0$, $\Pr[X_{t+1} = s_{t+1} \mid X_t = s_t, \ldots, X_0 = s_0] = \Pr[X_{t+1} = s_{t+1} \mid X_t = s_t]$. This so-called *Markov property* (sometimes called the memoryless property) implies that the probability of transitions to future states, such as s_{t+1}, depend only on the present state s_t, and so are independent of the remote past, namely $s_{t-1}, \ldots, s_0$ (Iosifescu, 1980).

A Markov chain is *stationary* iff $\forall\, t \text{ s.t. } \Pr[X_t = s_t] > 0$, $\Pr[X_{t+1} = s_{t+1} \mid X_t = s_t]$ remains constant. Since $\forall s \in \mathcal{S},\ \exists t_s \geq 0$ such that $\Pr\left[X_{t_s} = s\right] > 0$, given an enumeration, ι, of the state space, $\mathcal{S}$, there is an $n \times n$ matrix, M, such that $\Pr[X_{t+1} = \iota(i) \mid X_t = \iota(j)] = M_{i,j}$, whenever $\Pr[X_t = \iota(j)] > 0$. Notice that this implies that $n \geq |\mathcal{S}|$. When this holds, we say that M is a *transition matrix* of the chain *consistent with* ι. If $n = |\mathcal{S}|$, then M is uniquely defined, and we say that M is a *minimal* transition matrix of the chain.

Notice that if M_1 and M_2 are two minimal transition matrices, consistent with ι_1 and ι_2, respectively, then $M_2 = P^{-1} M_1 P$, where P is the permutation matrix such that $P_{i,j} = [\iota_1(i) = \iota_2(j)]$. In particular, there is a unique minimal, ι-consistent transition matrix for which ι is increasing. As we will see, using the following Lemma, one can show that, in a certain sense, the converse holds, as well.

Lemma 5.22. *For every sequence,* $\sigma \in S_n(k)$,

a) $\Pr[X_{t-k} = \iota(\sigma_0), \ldots X_t = \iota(\sigma_k)] \neq 0$ *iff* $\sigma \in \mathcal{P}_M(k)$ *and* $\Pr[X_{t-k} = \iota(\sigma_0)] \neq 0$.

b) *More specifically,*

$$\Pr[X_{t-k} = \iota(\sigma_0), \ldots X_t = \iota(\sigma_k)] = M_\sigma \Pr[X_{t-k} = \iota(\sigma_0)] \tag{5.6}$$

Proof. We prove both parts by induction on k. When $k = 0$, $M_\sigma = 1$ and $\sigma \in \mathcal{P}_M(0)$, so both parts are trivially true. In general, for any $\sigma \in S_n(k)$, take $\sigma' \in S_n(k-1)$ so that $\sigma'_i = \sigma_i$ for $0 \leq i < k$. We may then prove part b) in two cases.

If $\Pr[X_{t-k} = \iota(\sigma_0), \ldots X_{t-1} = \iota(\sigma_{k-1})] = 0$, then $\Pr[X_{t-k} = \iota(\sigma_0), \ldots X_t = \iota(\sigma_k)] = 0$. Moreover, by induction, we know that either $\Pr[X_{t-k} = \iota(\sigma_0)] = 0$, in which case Equation 5.6 is trivially satisfied, or $\sigma' \notin \mathcal{P}_M(k-1)$, in which case, $\sigma \notin \mathcal{P}_M(k)$, as well, so that $M_\sigma = 0$, and Equation 5.6 again holds.

Now assume that $\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] > 0$. The Markov property, along with the induction hypothesis, then implies that

$$\begin{aligned}
\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots X_t = \iota\left(\sigma_k\right)\right] &= \Pr\left[X_t = \iota\left(\sigma_k\right) \mid X_{t-k} = \iota\left(\sigma_0\right), \ldots X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&\quad \Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&= \Pr\left[X_t = \iota\left(\sigma_k\right) \mid X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&\quad M_{\sigma'} \Pr\left[X_{t-k} = \iota\left(\sigma_0'\right)\right] \\
&= M_{\sigma_k, \sigma_{k-1}} M_{\sigma'} \Pr\left[X_{t-k} = \iota\left(\sigma_0'\right)\right] \\
&= M_{\sigma} \Pr\left[X_{t-k} = \iota\left(\sigma_0\right)\right]
\end{aligned}$$

where we also appeal to the definition of M_σ and the fact that $\sigma_0 = \sigma_0'$. □

Obviously, the joint distribution of $\{X_t\}_{t=0}^k$, for any k, is determined by the joint distribution of X_*. Conversely, the sequence of all such joint distributions (i.e., for $k = 0, \ldots$) determine the joint distribution of the X_*. Lemma 5.22 says that, for a stationary Markov process, this sequence of joint distributions is equivalent, up to labelling of the states, ι, to an initial distribution (i.e., for X_0) and a transition matrix, M.

While ι allows us to associate states with indices, it is also helpful to associate states with the vertices of the standard, n-simplex, Δ_n. Specifically, let the state $\iota(i)$ correspond with the vertex, e_i of the standard, n-simplex, Δ_n. That is, given a chain, X_*, and an enumeration of its state space, i, we may define an associated, vector-valued chain, $\vec{X}_*$, where $\vec{X}_t(\omega) = e_j$, if $X_t(\omega) = \iota(j)$, and 0, otherwise. This form of the chain has the advantage that we may cleanly establish the connection between the probability distribution of X_t and the corresponding distribution vector.

Lemma 5.23. *Given a Markov chain, X_*, and an enumeration of its state space, ι, define the associated, vector-valued chain, $\vec{X}_*$, as above. For any t, $\mathbb{E}\left[\vec{X}_t\right]$ is then a vector with $\left(\mathbb{E}\left[\vec{X}_t\right]\right)_j = \Pr\left[X_t = \iota\left(j\right)\right]$ for all j.*

Proof. By definition, $\mathbb{E}\left[\vec{X}_t\right] = \sum_{k \in S_n} e_k \Pr\left[\vec{X}_t = e_k\right] = \sum_{j \in S_n} e_k \Pr\left[X_t = \iota\left(k\right)\right]$. Thus, $\left(\mathbb{E}\left[\vec{X}_t\right]\right)_j = e_j^t \mathbb{E}\left[\vec{X}_t\right] = \sum_{j \in S_n} e_j^t e_k \Pr\left[X_t = \iota\left(k\right)\right] = \Pr\left[X_t = \iota\left(j\right)\right]$. □

Alternatively, we may associate states with vertices in a directed graph. Specifically, we may view a stationary Markov chain with transition matrix M in terms of a random

walk on the weighted graph, $G(M)$, where the state, $\iota(i)$, corresponds to the vertex, v_i. Since $G(M)$ has no repeated edges, a walk of length N is equivalent to a sequence of adjacent vertices, which, by Lemma 5.22 a), corresponds to a possible sample from $\{X_t\}_{t=0}^N$. Moreover, if we choose the initial vertex according to the distribution of X_0 and the subsequent edges according to the edges weights, by Lemma 5.22 b), the probability of obtaining any given walk is the same as the probability of obtaining the corresponding sample from $\{X_t\}_{t=0}^N$. Thus, the graph, $G(M)$, and an initial distribution give an alternative, geometric characterization of the chain.

As before, we may carry over the terminology of strongly connected components, closed classes, invariant and transient sets of vertices in $G(M)$ from Section 1.1 and apply it to sets of states of a stationary Markov process. Notice M_σ is represents the conditional probability of the random walk realizing the specific sequence of states corresponding to σ, given that $X_0 = \iota(\sigma_0)$. Thus, a subset of states is invariant iff the probability of ever transitioning away from the set is 0. Likewise, any transient state has a positive probability of transitioning away from it without ever returning.

Given a chain, X_*, and a Borel set, $\beta \in \mathcal{B}$, we want to define a new chain, $\tilde{X}_*$, where we "collapse" the time spent in β. To make this precise, define the functions $\tau_{\beta,k} : \Omega \to \mathbb{R}$ inductively, as follows. Set $\tau_{\beta,-1}(\omega) = -1$, and define

$$\tau_{\beta,k}(\omega) = \min\left\{t > \tau_{\beta,k-1}(\omega) \mid X_t(\omega) \notin \beta\right\}$$

with the convention that $\min \emptyset = \infty$.

Lemma 5.24. *If $\tau_{\beta,k'}(\omega) < \infty$, for all $0 \le k \le k'$,*

$$\tau_{\beta,k}(\omega) = \min\left\{t \ge 0 \mid k+1 = \left|\left\{0 \le t' \le t \mid X_{t'}(\omega) \in \overline{\beta}\right\}\right|\right\} \tag{5.7}$$

Proof. By definition, since $\tau_{\beta,k'}(\omega) < \infty$, $\tau_{\beta,k'-1}(\omega) < \tau_{\beta,k'}(\omega)$, so that $\tau_{\beta,k'-1}(\omega) < \infty$, and so on. In particular, $\left\{\tau_{\beta,k}(\omega)\right\}_{k=0}^{k'}$ is an increasing sequence. For the remainder of the proof, we will drop the notation for evaluation, since all random variables (i.e., X_t's and $\tau_{\beta,k}$'s) will always be evaluated at a fixed value, ω. Likewise, we will always assume that $0 \le k \le k'$.

Now define $T_t = \left\{0 \le t' \le t \mid X_{t'} \in \overline{\beta}\right\}$ and $T'_t = \left\{t' > t \mid X_{t'} \in \overline{\beta}\right\}$. With this notation, $\tau_{\beta,k} = \min T'_{\tau_{\beta,k-1}}$. We now show that

$$T'_{\tau_{\beta,k-1}} \cap T_{\tau_{\beta,k}} = \left\{\tau_{\beta,k}\right\} \tag{5.8}$$

Since $\tau_{\beta,k} \in T'_{\tau_{\beta,k-1}}$, we must have $\tau_{\beta,k} \in T_{\tau_{\beta,k}}$. Moreover, this is the only value in $T'_{\tau_{\beta,k-1}} \cap T_{\tau_{\beta,k}}$. If t were a different value in the intersection, since $\tau_{\beta,k}$ is minimum in $T'_{\tau_{\beta,k-1}}$, we must have $\tau_{\beta,k} < t$. But since $t \in T_{\tau_{\beta,k}}$, we should have $t \leq \tau_{\beta,k}$, a contradiction.

Now define $\delta(t) \equiv |T_t|$. Now observe that $|T_t| = |T_{t-1}| + \left[X_t \in \overline{\beta}\right]$, so that $\delta(t) \equiv |T_t|$ is non-decreasing. Using this notation, Equation 5.7 may be rewritten as $\tau_{\beta,k} = \min \delta^{-1}(\{k+1\})$. In particular, we must show that $\delta\left(\tau_{\beta,k}\right) = k+1$.

Since the $\tau_{\beta,j}$'s are increasing for $0 \leq j \leq k'$, we have that

$$\begin{aligned} T_{\tau_{\beta,k}} &= T_{\tau_{\beta,k}} \cap \left(\tau_{\beta,-1}, \tau_{\beta,k}\right] = T_{\tau_{\beta,k}} \cap \bigcup_{j=0}^{k} \left(\tau_{\beta,j-1}, \tau_{\beta,j}\right] \\ &= \bigcup_{j=0}^{k} T_{\tau_{\beta,k}} \cap \left(\tau_{\beta,j-1}, \tau_{\beta,j}\right] = \bigcup_{j=0}^{k} T'_{\tau_{\beta,j-1}} \cap T_{\tau_{\beta,j}} = \bigcup_{j=0}^{k} \left\{\tau_{\beta,j}\right\} \end{aligned}$$

Therefore, since the $\tau_{\beta,j}$'s are distinct,

$$\delta\left(\tau_{\beta,k}\right) = \left|T_{\tau_{\beta,k}}\right| = \left|\bigcup_{j=0}^{k} \left\{\tau_{\beta,j}\right\}\right| = k+1$$

Thus, $\tau_{\beta,k} \in \delta^{-1}(\{k+1\})$.

Now if $\tau_{\beta,k}$ is not minimum, then there is some $t < \tau_{\beta,k}$ such that $\delta(t) = k+1$. Since, $\delta\left(\tau_{\beta,k-1}\right) = k$ and δ is increasing, we must have $\tau_{\beta,k-1} < t < \tau_{\beta,k}$. As before,

$$\begin{aligned} T_t &= T_t \cap \left(\tau_{\beta,-1}, \tau_{\beta,k}\right] = T_t \cap \bigcup_{j=0}^{k} \left(\tau_{\beta,j-1}, \tau_{\beta,j}\right] \\ &= \bigcup_{j=0}^{k} T_t \cap \left(\tau_{\beta,j-1}, \tau_{\beta,j}\right] = T'_{\tau_{\beta,k-1}} \cap T_t \cup \bigcup_{j=0}^{k-1} T'_{\tau_{\beta,j-1}} \cap T_{\tau_{\beta,j}} \\ &= T'_{\tau_{\beta,k-1}} \cap T_t \cup \bigcup_{j=0}^{k-1} \left\{\tau_{\beta,j}\right\} \end{aligned}$$

Since $\delta(t) = |T_t| = k+1$, we must have $\left|T'_{\tau_{\beta,k-1}} \cap T_t\right| = 1$; in particular, there is some $t' \in T'_{\tau_{\beta,k-1}} \cap T_t$.

Since $t' \leq t < \tau_{\beta,k}$, $T_t \subset T_{\tau_{\beta,k}}$. Therefore, $t' \in T'_{\tau_{\beta,k-1}} \cap T_t \subset T'_{\tau_{\beta,k-1}} \cap T_{\tau_{\beta,k}} = \left\{\tau_{\beta,j}\right\}$, so that $t' = \tau_{\beta,k}$, a contradiction.

Thus, $\tau_{\beta,k} = \min \delta^{-1}(\{k+1\})$. Since $k \in \left[0, k'\right]$ was arbitrary, we have proven Equation 5.7 for all $0 \leq k \leq k'$. □

Using the language of Markov chain theory, $\tau_{\beta,k}$ is the $k+1^{\text{st}}$ "hitting time" for $\overline{\beta}$. Notice that

$$\left\{\tau_{\beta,k}=t\right\}=\begin{cases}\left\{X_t\in\overline{\beta}\right\}\cap\left\{\tau_{\beta,k-1}=t'\right\}\cap\bigcap_{t'<t''<t}\left\{X_{t''}\in\beta\right\} & \text{if } t<\infty\\ \left\{\tau_{\beta,k-1}=t'\right\}\cap\bigcap_{t'<t''}\left\{X_{t''}\in\beta\right\} & \text{if } t=\infty\end{cases}$$

Therefore, by induction, this event be expressed solely in terms of $X_0,\ldots,X_t$. In this case, we say $\tau_{\beta,k}$ is a *Markov time* (Iosifescu, 1980).

We now prove that $\tau_{\beta,k}$ is almost always finite, when β is open. In fact, this effectively characterizes open sets of states.

Lemma 5.25. *A set of states, β, of a finite-state, stationary Markov process, X_*, is open iff $\Pr\left[\bigcap_{r\geq j}X_r^{-1}(\beta)\right]=0,\forall j$. In particular, if β is open, $\tau_{\beta,0}<\tau_{\beta,1}<\cdots<\infty$ with probability 1.*

Proof. Assume that the process is ι-consistent with a matrix, M, with state space, $\mathcal{S}$, and let $s=\iota^{-1}(\beta)$. Then, using the notation of Theorem 4.5, the u^{th} column sum of $\overline{M}^q$ is

$$J\,\overline{M}^q e_u=\sum_{v=1}^{|s|}\left(\overline{M}^q\right)_{v,u}=\sum_{v\in S_{|s|}}\sum_{\sigma\in S_{|s|}(v,u,q)}\overline{M}_\sigma=\sum_{v\in S_{|s|}}\sum_{\sigma\in S_{|s|}(v,u,q)}M_{s_\sigma}\tag{5.9}$$

Therefore,

$$\begin{aligned}\Pr\left[\bigcap_{r=j}^{j+q}X_r^{-1}(\beta)\right] &= \Pr\left[X_j\in\beta,\ldots,X_{j+q}\in\beta\right]=\Pr\left[X_j\in\beta\cap\mathcal{S},\ldots,X_{j+q}\in\beta\cap\mathcal{S}\right]\\ &= \sum_{i_0,\ldots,i_q\in s}\Pr\left[X_j=\iota(i_0),\ldots X_{j+q}=\iota\left(i_q\right)\right]\\ &= \sum_{\sigma\in S_{|s|}(q)}\Pr\left[X_j=\iota\left(s_{\sigma_0}\right),\ldots X_{j+q}=\iota\left(s_{\sigma_q}\right)\right]\end{aligned}$$

and

$$\begin{aligned}\Pr\left[\bigcap_{r=j}^{j+q}X_r^{-1}(\beta)\right] &= \sum_{\sigma\in S_{|s|}(q)}M_{s_\sigma}\Pr\left[X_j=\iota\left(s_{\sigma_0}\right)\right] && \text{by Equation 5.6}\\ &= \sum_{u,v\in S_{|s|}}\sum_{\sigma\in S_{|s|}(v,u,q)}M_{s_\sigma}\Pr\left[X_j=\iota(s_u)\right]\\ &= \sum_{u\in S_{|s|}}J\,\overline{M}^q e_u\Pr\left[X_j=\iota(s_u)\right] && \text{by Equation 5.9}\\ &= J\,\overline{M}^q\sum_{u\in S_{|s|}}e_u\Pr\left[X_j=\iota(s_u)\right]\end{aligned}$$

Thus,

$$\Pr\left[\bigcap_{r\geq j} X_r^{-1}(\beta)\right] = \lim_{q\to\infty} \Pr\left[\bigcap_{r=j}^{j+q} X_r^{-1}(\beta)\right] = \lim_{q\to\infty} J\,\overline{M}^q \sum_{u\in S_{|s|}} e_u \Pr\left[X_j = \iota(s_u)\right]$$

If β is open, by Theorem 4.5 a), $\lim_{q\to\infty} J\,\overline{M}^q = 0$ so that $\Pr\left[\bigcap_{r\geq j} X_r^{-1}(\beta)\right] = 0$.

Conversely, assume that $\Pr\left[\bigcap_{r\geq j} X_r^{-1}(\beta)\right] = 0, \forall\, j$. For any u', since $\iota(s_{u'}) \in \mathcal{S}$, there is some j for which $\Pr\left[X_j = \iota(s_{u'})\right] > 0$. Moreover,

$$\Pr\left[\bigcap_{r=j}^{j+q} X_r^{-1}(\beta)\right] = J\,\overline{M}^q \sum_{u\in S_{|s|}} e_u \Pr\left[X_j = \iota(s_u)\right] \geq J\,\overline{M}^q e_{u'} \Pr\left[X_j = \iota(s_{u'})\right] \geq 0$$

Since the left-hand side goes to 0 as $q \to \infty$, and $\Pr\left[X_j = \iota(s_{u'})\right] \neq 0$, we must have $\lim_{q\to\infty} J\,\overline{M}^q e_{u'}$. Since u' was arbitrary, $0 = J\left(\lim_{q\to\infty} \overline{M}^q\right) J^t = 0$, so that $\lim_{q\to\infty} \overline{M}^q = 0$. Appealing to the contrapositive of Theorem 4.5 d), we may conclude that β is open.

By definition, $\tau_{\beta,t}(\omega) \leq \tau_{\beta,t+1}(\omega)$, with inequality, unless $\tau_{\beta,t}(\omega) = \infty$. However, $\tau_{\beta,t}(\omega) = \infty$ iff $\left\{t' > \tau_{\beta,t-1}(\omega) \mid X_{t'}(\omega) \notin \beta\right\} = \emptyset$, that is, $X_{t'}(\omega) \in \beta, \forall\, t' > \tau_{\beta,t-1}(\omega)$. Thus, $\tau_{\beta,0} < \tau_{\beta,1} < \cdots < \infty$, outside of the set $\bigcup_j \bigcap_{r\geq j} X_r^{-1}(\beta)$. But we have just shown that, when β is open,

$$\Pr\left[\bigcup_j \bigcap_{r\geq j} X_r^{-1}(\beta)\right] = \sum_j \Pr\left[\bigcap_{r\geq j} X_r^{-1}(\beta)\right] = \sum_j 0 = 0$$

That is, $\tau_{\beta,0} < \tau_{\beta,1} < \cdots < \infty$ occurs with probability 1. □

By Lemma 5.25, $\tau_{\beta,t}$ is a Markov time with $\Pr[\tau_{\beta,t} < \infty] = 1$. Such a random variable is known as a *stopping time*. Evaluating a Markov chain at a stopping time is also a random variable (Iosifescu, 1980). Thus, if β is open, we may define $\pi_{\beta,t}(X_*) \equiv X_{\tau_{\beta,t}}$, where we define $\pi_{\beta,t}(X_*) = \min\overline{\beta}$, when $\tau_{\beta,t} = \infty$. In this way, we have defined the desired chain, $\tilde{X}_* \equiv \pi_{\beta,*}(X_*)$. We will show that $\pi_{\beta,*}$ is an operator on Markov chains which corresponds directly to applying the reduction construction of section 5.2 to the transition matrix of the chain. Notice that, as we mentioned in section 5.2, from a Markov chain point of view, we have simply reduced the time spent in the states of β to 0.

We now wish to identify the transition matrix for $\pi_{\beta,*}(X_*)$. The proof will be similar to that of Theorem 5.8, but we will need to generalize the notation from Chapter 1 a bit. Define

$$\mathcal{S}_n(s,i,j,l,m) = \{\sigma \in \mathcal{S}_n(i,j,l+m) \mid m = |\{k \in (0,l+m) \mid \sigma_k \notin s\}|\}$$

i.e., sequences of length $l+m+1$ starting at j, ending at i, whose interior values lie outside of s exactly m times. Thus, for example, $\mathcal{S}_n(s,i,j,l,0) = \mathcal{S}_n(s,i,j,l)$.

Using this notation, we may now identify the distribution of $\widetilde{X}_t = \pi_{\beta,t}(X_*)$. That is, we can give a formula for $\Pr\left[\widetilde{X}_t = x\right]$.

Theorem 5.26. *If X_* is a finite state, stationary Markov chain with state space, $\mathcal{S}$, which is ι-consistent with a minimal, $n\times n$ transition matrix, M, β is open, and $\widetilde{X}_* = \pi_{\beta,*}(X_*)$, then $\overline{\beta}$ is the state space for $\widetilde{X}_*$, so that $\iota'(k) = \iota(\overline{s}_k)$ enumerates $\overline{\beta}$, where $s = \iota^{-1}(\beta) \subset S_n$. For $t > 0$,*

$$\begin{aligned}\Pr\left[\widetilde{X}_t = \iota'(k)\right] &= \sum_{l=0}^{\infty}\sum_{j\in s}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \\ &\quad + \sum_{l=0}^{\infty}\sum_{j\in \overline{s}}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t-1)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \qquad (5.10)\end{aligned}$$

and

$$\Pr\left[\widetilde{X}_0 = \iota'(k)\right] = \sum_j p_{k,j}\Pr\left[X_0 = \iota(j)\right] \qquad (5.11)$$

where $(\widehat{M}, p, \imath)$ is the reduction of M with respect to s. That is, the distribution of $\widetilde{X}_0$, is result of applying the projection, p, to the distribution of X_0, where we view each distribution as an column vector, as in Lemma 5.23.

Proof. By definition, $\pi_{\beta,t}(X_*)(\omega) \in \overline{\beta}$ for all t and ω. Notice also that, since $\iota(s_*)$ enumerates $\beta\cap\mathcal{S}$, $\iota' = \iota(\overline{s}_*)$ enumerates $\overline{\beta}$. In particular, the state space for $\widetilde{X}_*$ is contained in $\overline{\beta}$ and $\widetilde{X}_t = x$ iff $x = \iota(\overline{s}_k)$ for some $k \in S_{|\overline{s}|}$.

Now remember that, when $\tau_{\beta,t} < \infty$, $\left\{\tau_{\beta,j}\right\}_{j=-1}^{t}$ is an increasing sequence of integers.

In particular, $\tau_{\beta,t} \geq t$. Thus, by Lemma 5.25,

$$\begin{aligned}
&\Pr\left[\widetilde{X}_t = \iota'(k)\right] = \Pr\left[\widetilde{X}_t = \iota'(k), \tau_{\beta,t} < \infty\right] = \\
&= \Pr\left[X_{\tau_{\beta,t}} = \iota(\overline{s}_k)\right] = \sum_{l=0}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \\
&= [t=0] \Pr\left[X_t = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] + \sum_{l=[t=0]}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \\
&= [t=0] \Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] + \sum_{l=[t=0]}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \\
&= [t=0] \Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] + \sum_{l=[t=0]}^{\infty} \sum_{j=1}^{n} \Pr\left[X_{l+t} = \iota(\overline{s}_k), X_0 = \iota(j), l = \tau_{\beta,t} - t\right]
\end{aligned}$$

where divided into cases, based on $l = \tau_{\beta,t} - t$ and the initial state of the process, and we used a notational trick to pull out the $l = 0$ term from the summation, when $t = 0$. Notice that, for all the terms in the summation, we are guaranteed that $l + t > 0$.

Assuming that $X_0 = \iota(j)$, $X_{l+t} = \iota(\overline{s}_k)$, $l = \tau_{\beta,t} - t$, and $l + t > 0$ (i.e., in the context of any of the summation terms), by Lemma 5.24, $\{X_i\}_{i=0}^{l+t}$ takes exactly $t + 1$ values in $\overline{\beta}$ and l values in β. In this case, we may write $X_i = \iota(\sigma_i)$ for $0 \leq i \leq l + t$, for some

$\sigma \in S_n(s, \overline{s}_k, j, l, t-1)$, if $j \notin s$, and $\sigma \in S_n(s, \overline{s}_k, j, l, t)$, if $j \in s$. Thus,

$$\begin{aligned}
\Pr\left[\widetilde{X}_t = \iota'(k)\right] &= [t=0]\Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in \overline{s}} \Pr\left[X_{l+t} = \iota(\overline{s}_k), X_0 = \iota(j), \text{ and } l = \tau_{\beta,t} - t\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in s} \Pr\left[X_{l+t} = \iota(\overline{s}_k), X_0 = \iota(j), \text{ and } l = \tau_{\beta,t} - t\right] \\
&= [t=0]\Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in s} \sum_{\sigma \in S_n(s,\overline{s}_k,j,l,t)} \Pr\left[X_i = \iota(\sigma_i), 0 \le i \le l+t\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in \overline{s}} \sum_{\sigma \in S_n(s,\overline{s}_k,j,l,t-1)} \Pr\left[X_i = \iota(\sigma_i), 0 \le i \le l+t\right] \\
&= [t=0]\Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in s} \sum_{\sigma \in S_n(s,\overline{s}_k,j,l,t)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \\
&\quad + \sum_{l=[t=0]}^{\infty} \sum_{j \in \overline{s}} \sum_{\sigma \in S_n(s,\overline{s}_k,j,l,t-1)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \qquad (5.12)
\end{aligned}$$

where we appeal to Lemma 5.22 b) for the final step.

Equation 5.12 simplifies to Equation 5.10, when $t > 0$. Moreover, it implies that the state space for $\widetilde{X}_*$ equals all of $\overline{\beta}$, because, for any state $\iota'(k) \in \overline{\beta} \subset \mathcal{S}$, there is some t for which $\Pr\left[X_t = \iota'(k)\right] > 0$. If $t = 0$, the first term in Equation 5.12 is non-zero, so that $\Pr\left[\widetilde{X}_t = \iota'(k)\right] > 0$.

Otherwise, by Lemma 5.22, and the fact that ι enumerates the state space of X_*,

$$\begin{aligned}
\Pr\left[X_t = \iota'(k)\right] &= \sum_{j \in S_n} \Pr\left[X_t = \iota(\overline{s}_k), X_0 = \iota(j)\right] \\
&= \sum_{j \in S_n} \sum_{\sigma \in S_n(\overline{s}_k,j,t)} \Pr\left[X_i = \iota(\sigma_i), i = 0, \ldots, t\right] = \sum_{j \in S_n} \sum_{\sigma \in S_n(\overline{s}_k,j,t)} M_\sigma \Pr\left[X_0 = \iota(j)\right]
\end{aligned}$$

Since this is non-zero, it must have at least one non-zero term, corresponding to some $j \in S_n$ and $\sigma \in S_n(\overline{s}_k, j, t)$, such that $0 < \Pr\left[X_0 = \iota(j)\right], M_\sigma$. If $l+1$ is the number of values of σ in $\overline{s}$, then either $\sigma \in S_n(s, \overline{s}_k, j, l, t)$, if $j \in s$, or $\sigma \in S_n(s, \overline{s}_k, j, l, t-1)$, otherwise. In any case, Equation 5.10 has at least one non-zero term, so that $\Pr\left[\widetilde{X}_t = \iota'(k)\right] > 0$. Therefore, $\overline{\beta}$ is the state space for $\widetilde{X}_*$.

To prove Equation 5.11, take $t = 0$. Since $S_n(s, \overline{s}_k, j, l, t-1) = \emptyset$, Equation 5.12 simplifies to

$$\begin{aligned}
\Pr\left[\widetilde{X}_0 = \iota'(k)\right] &= \Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right] + \sum_{l=1}^{\infty}\sum_{j\in s}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l,0)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \\
&= \Pr\left[X_0 = \iota(\overline{s}_k)\right] + \sum_{l=1}^{\infty}\sum_{j\in s}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \\
&= \Pr\left[X_0 = \iota(\overline{s}_k)\right] + \sum_{l=0}^{\infty}\sum_{j\in s}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l+1)} M_\sigma \Pr\left[X_0 = \iota(j)\right] \\
&= \Pr\left[X_0 = \iota(\overline{s}_k)\right] + \sum_{l=0}^{\infty}\sum_{j\in S_{|s|}}\sum_{\sigma\in S_n(s,\overline{s}_k,s_j,l+1)} M_\sigma \Pr\left[X_0 = \iota\left(s_j\right)\right] \\
&= \Pr\left[X_0 = \iota(\overline{s}_k)\right] + \sum_{j\in S_{|s|}}\sum_{l=0}^{\infty}\sum_{\sigma\in S_n(s,\overline{s}_k,s_j,l+1)} M_\sigma \Pr\left[X_0 = \iota\left(s_j\right)\right] \quad (5.13)
\end{aligned}$$

Now observe that from Theorem 5.6 and Theorem 4.5 b),

$$\begin{aligned}
p &= \begin{pmatrix} I & -\overline{N}\,\overline{\Lambda}^{-1} \end{pmatrix} P_s^t = \begin{pmatrix} I & -\overline{N}\,\overline{\Lambda}^{-1} \end{pmatrix} \begin{pmatrix} \pi_{\overline{s}} \\ \pi_s \end{pmatrix} \\
&= \pi_{\overline{s}} - \overline{N}\,\overline{\Lambda}^{-1}\pi_s = \pi_{\overline{s}} + \sum_{l=0}^{\infty} \overline{N}\,\overline{M}^l \pi_s
\end{aligned}$$

Thus, using Lemma 2.7 a), we have

$$\begin{aligned}
\sum_{j\in S_n} p_{k,j} \Pr\left[X_0 = \iota(j)\right] &= \sum_{j\in\overline{s}} p_{k,j} \Pr\left[X_0 = \iota(j)\right] + \sum_{j\in s} p_{k,j} \Pr\left[X_0 = \iota(j)\right] \\
&= \sum_{j\in\overline{s}} e_k^t p e_j \Pr\left[X_0 = \iota(j)\right] + \sum_{j\in s} e_k^t p e_j \Pr\left[X_0 = \iota(j)\right] \\
&= \sum_{j\in\overline{s}} e_k^t \pi_{\overline{s}} e_j \Pr\left[X_0 = \iota(j)\right] + \sum_{j\in s} e_k^t \sum_{l=0}^{\infty} \overline{N}\,\overline{M}^l \pi_s e_j \Pr\left[X_0 = \iota(j)\right] \\
&= \Pr\left[X_0 = \iota(\overline{s}_k)\right] + \sum_{j\in S_{|s|}}\sum_{l=0}^{\infty} e_k^t \overline{N}\,\overline{M}^l e_j \Pr\left[X_0 = \iota\left(s_j\right)\right] \quad (5.14)
\end{aligned}$$

Comparing Equations 5.13 and 5.14, we see that it only remains to show that $e_k^t \overline{N}\,\overline{M}^l e_j$

equals $\sum_{\sigma\in S_n(s,\overline{s}_k,s_j,l+1)} M_\sigma$. As in the proof of Theorem 5.8,

$$\begin{aligned} e_k^t \overline{N}\,\overline{M}^l e_j &= \sum_{j'\in S_{|s|}} e_k^t \overline{N} e_{j'} e_{j'} \overline{M}^l e_j = \sum_{j'\in S_{|s|}} M_{\overline{s}_k, s_{j'}} \sum_{\sigma'\in\mathcal{S}_n\left(s,s_j,s_{j'},l\right)} M_{\sigma'} \\ &= \sum_{j'\in S_{|s|}} \sum_{\sigma'\in\mathcal{S}_n\left(s,s_j,s_{j'},l\right)} M_{\overline{s}_k,s_{j'}} M_{\sigma'} = \sum_{\sigma\in S_n(s,\overline{s}_k,s_j,l+1)} M_\sigma \end{aligned}$$

□

We may also identify a transition matrix for $\widetilde{X}_* = \pi_{\beta,*}(X_*)$. Intuitively, the following theorem says that the transition matrix for $\widetilde{X}_*$ is the result of applying the reduction construction of section 5.2 to the transition matrix for X_*. The proof is similar to that of Theorem 5.26, but it will be helpful to alter our notation a bit. Define

$$\mathcal{S}'_n(s,i,l,m) = \{\sigma\in\mathcal{S}_n(l+m) \mid \sigma_{l+m} = i, m = |\{k < l+m) \mid \sigma_k \notin s\}|\}$$

i.e., sequences of length $l+m+1$, ending at i, whose values, excepting the last, lie outside of s exactly m times. Notice that this time we do not specify the initial value and we do not exclude it from our count of values in $\overline{s}$.

Theorem 5.27. *If X_* is a finite state, stationary Markov chain which is ι-consistent with transition matrix, M, and β is open, then $\pi_{\beta,*}(X_*)$ is a stationary Markov chain ι'-consistent with transition matrix, $\widehat{M}$, where $\widehat{M}$ is the reduction of M with respect to $s = \iota^{-1}(\beta)$ and $\iota'(j) = \iota(\overline{s}_j)$.*

Proof. The proof is similar to that of Theorem 5.26. As before, define $\widetilde{X}_t = \pi_{\beta,t}(X_*)$, and recall that $\iota(s_*)$ and $\iota(\overline{s}_*)$ enumerate $\beta\cap\mathcal{S}$ and $\overline{\beta}$, respectively. By Lemma 5.25, we may restrict attention to the case, $\tau_{\beta,t} < \tau_{\beta,t+1} < \infty$ and again exploit the fact that $t\le\tau_{\beta,t}$.

$$\begin{aligned} &\Pr\left[\widetilde{X}_{t+1} = \iota'(k), \widetilde{X}_t = \iota'\left(k'\right)\right] = \Pr\left[\widetilde{X}_{t+1} = \iota'(k), \widetilde{X}_t = \iota'\left(k'\right), \tau_{\beta,t+1} < \infty\right] = \\ &= \sum_{m=0}^{\infty}\sum_{l=0}^{\infty} \Pr\left[X_{l+m+t+1} = \iota(\overline{s}_k), X_{m+t} = \iota(\overline{s}_{k'}), l = \tau_{\beta,t+1} - m - t - 1, m = \tau_{\beta,t} - t\right] \end{aligned}$$

Assuming that $X_{m+t} = \iota(\overline{s}_{k'})$ and $m = \tau_{\beta,t} - t$, Lemma 5.24 implies that $X_{l+m+t+1} = \iota(\overline{s}_k)$ with $l = \tau_{\beta,t+1} - m - t - 1$ iff $X_i\in\beta$ for $m+t < i \le l+m+t$ with $X_{l+m+t+1} = \iota(\overline{s}_k)$.

Thus,

$$\Pr\left[\widetilde{X}_{t+1}=\iota'(k),\widetilde{X}_t=\iota'\left(k'\right)\right]=$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}\Pr\left[X_{m+t+i}=\iota\left(\sigma'_i\right),i=0,\ldots,l+1,m=\tau_{\beta,t}-t\right]$$

Appealing again to Lemma 5.24, we see that $X_{m+t}=\iota\left(\overline{s}_{k'}\right)$ and $m=\tau_{\beta,t}-t$ iff $\{X_i\}_{i=0}^{m+t-1}$ takes exactly t values in $\overline{\beta}$, i.e., we may write $X_i=\iota\left(\sigma''_i\right)$ for $0\leq i\leq m+t$, for some $\sigma''\in S'_n(s,\overline{s}_k,m,t)$. Thus, for $\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)$,

$$\Pr\left[X_{m+t+i}=\iota\left(\sigma'_i\right),i=0,\ldots,l+1,m=\tau_{\beta,t}-t\right]=$$
$$=\sum_{\sigma''\in S'_n(s,\overline{s}_{k'},m,t)}\Pr\left[X_{m+t+i}=\iota\left(\sigma'_i\right),i=0,\ldots,l+1,X_i=\iota\left(\sigma''_i\right),i=0,\ldots,m+t\right]$$

Therefore,

$$\Pr\left[\widetilde{X}_{t+1}=\iota'(k),\widetilde{X}_t=\iota'\left(k'\right)\right]=$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}\sum_{\sigma''\in S'_n(s,\overline{s}_{k'},m,t)}\Pr\left[X_i=\iota(\sigma_i),i=0,\ldots,l+t,\sigma=\sigma'*\sigma''\right]$$

where $\sigma'*\sigma''$ is the concatenation of the walks given by σ' and σ''. Specifically, $\sigma_i=\sigma''_i$, for $0\leq i\leq m+t$ and $\sigma_{i+m+t}=\sigma'_i$, for $0\leq i\leq l+1$.[3]

Now appealing to Lemma 5.22 b), we have

$$\Pr\left[\widetilde{X}_{t+1}=\iota'(k),\widetilde{X}_t=\iota'\left(k'\right)\right]=$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}\sum_{\sigma''\in S'_n(s,\overline{s}_{k'},m,t)}M_{\sigma'*\sigma''}\Pr\left[X_0=\iota\left(\sigma''_0\right)\right]$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}\sum_{\sigma''\in S'_n(s,\overline{s}_{k'},m,t)}M_{\sigma'}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma''_0\right)\right]$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}M_{\sigma'}\sum_{\sigma''\in S'_n(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma''_0\right)\right]$$

[3]Notice that we concatenate as *walks*, rather than as sequences; for example $(3,1,4)*(4,7,2)=(3,1,4,7,2)$, instead of $(3,1,4,4,7,2)$.

We may then exploit Theorem 5.8 as follows,

$$\begin{aligned}
&\Pr\left[\widetilde{X}_{t+1}=\iota'(k),\widetilde{X}_t=\iota'\left(k'\right)\right]=\\
&=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'},l+1\right)}M_{\sigma'}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]\\
&=\sum_{m=0}^{\infty}\sum_{\sigma'\in\mathcal{S}_n\left(s,\overline{s}_k,\overline{s}_{k'}\right)}M_{\sigma'}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]\\
&=\sum_{m=0}^{\infty}\sum_{\sigma'\in\mathcal{P}_M\left(s,\overline{s}_k,\overline{s}_{k'}\right)}M_{\sigma'}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]\\
&=\sum_{m=0}^{\infty}\widehat{M}_{k,k'}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]\\
&=\widehat{M}_{k,k'}\sum_{m=0}^{\infty}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]
\end{aligned}$$

Reversing our previous calculations gives

$$\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{\sigma''\in S_n'(s,\overline{s}_{k'},m,t)}M_{\sigma''}\Pr\left[X_0=\iota\left(\sigma_0''\right)\right]=\\
&\quad=\sum_{m=0}^{\infty}\Pr\left[X_{m+t}=\iota'\left(k'\right),m=\tau_{\beta,t}-t\right]=\Pr\left[\widetilde{X}_t=\iota'\left(k'\right)\right]
\end{aligned}$$

Therefore,

$$\Pr\left[\widetilde{X}_{t+1}=\iota'(k),\widetilde{X}_t=\iota'\left(k'\right)\right]=\widehat{M}_{k,k'}\Pr\left[\widetilde{X}_t=\iota'\left(k'\right)\right]$$

so that, $\Pr\left[\widetilde{X}_{t+1}=\iota'(k)\mid\widetilde{X}_t=\iota'\left(k'\right)\right]=\widehat{M}_{k,k'}$, when $\Pr\left[\widetilde{X}_t=\iota'\left(k'\right)\right]>0$. In particular X_* is a stationary Markov chain ι'-consistent with transition matrix, $\widehat{M}$. □

We now give an alternative description of the sequence, $\pi_{\beta,t}(X_*)(\omega)$, which applies almost always.

Lemma 5.28. *Given a Markov chain, X_*, and an open Borel set, β,*

$$\Pr\left[\omega\in\Omega\mid\tau_{\beta,k}(\omega)\textit{ for }k\geq0\textit{ is an increasing enumeration of }\left\{t\geq0\mid X_t(\omega)\in\overline{\beta}\right\}\right]=1$$

Proof. Define $T(\omega) \equiv \left\{t \geq 0 \mid X_t(\omega) \in \overline{\beta}\right\}$. We must show that, for $k \geq 0$, $\tau_{\beta,k}(\omega)$ is an increasing enumeration of $T(\omega)$, except on a set of probability 0. However, by definition, $\{\tau_{\beta,k}(\omega)\}_{k=-1}^{\infty}$ is a non-decreasing sequence starting at -1. Moreover, whenever $\tau_{\beta,k}(\omega)$ is finite, and $k \neq -1$, it is in T. By Lemma 5.25, it is strictly increasing and finite with probability 1.

Thus, choose an $\omega \in \Omega$ for which this is the case. For convenience, in the remainder of this proof, we will omit the notation for evaluation by ω, since everything can be assumed to be evaluated at ω. It remains to prove that, if $t \in T$, then $t = \tau_{\beta,k}$ for some $k \geq 0$. We may prove this by considering the contrapositive. If for all $k \geq 0$, $t \neq \tau_{\beta,k}$, there must be some $k \geq 0$ such that $\tau_{\beta,k-1} < t < \tau_{\beta,k}$. By definition, since $\tau_{\beta,k}$ is defined as a minimum value, we must have $X_t \in \beta$, that is, $t \notin T$. □

In other words, with probability 1, $\pi_{\beta,*}(X_*)$ is the result of deleting those entries of X_* with values in β.

Lemma 5.28 implies that this reduction operator is "natural" in the sense that it behaves as expected under iteration. Intuitively, if we first delete entries from a sequence with values in β_2 and then delete from the remaining entries those with values in β_2, we get the same result as if we had simply deleted those entries with values in $\beta_1 \cup \beta_2$.

Theorem 5.29. *Given a Markov chain, X_*, and open Borel sets, $\beta = \beta_1 \cup \beta_2$, $\pi_{\beta,*}(X_*) = \pi_{\beta_1,*}\left(\pi_{\beta_2,*}(X_*)\right)$ with probability 1.*

Proof. Let $X_*^2 = \pi_{\beta_2,*}(X_*)$, $X_*^1 = \pi_{\beta_1,*}\left(X_*^2\right)$, and $\tilde{X}_* = \pi_{\beta,*}(X_*)$. Likewise, let $T_2 = \left\{t \geq 0 \mid X_t \in \overline{\beta_2}\right\}$, $T_1 = \left\{t \geq 0 \mid X_t^2 \in \overline{\beta_1}\right\}$, and $T = \left\{t \geq 0 \mid X_t \in \overline{\beta}\right\}$. For $k \geq 0$, we may assume that $\tau_{\beta_2,k}$, $\tau_{\beta_1,k}$, and $\tau_{\beta,k}$ are increasing enumerations of T_2, T_1, and T, respectively, since, by Lemma 5.28, this holds except on a set of probability 0.

Now notice that

$$\begin{aligned}
\exists\, k \geq 0,\, t = \tau_{\beta_2,\tau_{\beta_1,k}} &\iff \exists\, k, t' \geq 0,\, t' = \tau_{\beta_1,k} \text{ and } t = \tau_{\beta_2,t'} \\
&\iff \exists\, t' \in T_1,\, t = \tau_{\beta_2,t'} && \text{since } \tau_{\beta_1,*} \text{ enumerates } T_1 \\
&\iff \exists\, t' \geq 0,\, X^2_{t'} \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2,t'} && \text{by definition of } T_1 \\
&\iff \exists\, t' \geq 0,\, X_{\tau_{\beta_2,t'}} \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2,t'} && \text{by definition of } X^2_* \\
&\iff \exists\, t' \geq 0,\, X_t \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2,t'} \\
&\iff X_t \in \overline{\beta_1} \text{ and } t \in T_2 && \text{since } \tau_{\beta_2,*} \text{ enumerates } T_2 \\
&\iff X_t \in \overline{\beta_1} \text{ and } X_t \in \overline{\beta_2} \\
&\iff X_t \in \overline{\beta_1} \cap \overline{\beta_2} = \overline{\beta_1 \cup \beta_2} = \overline{\beta} \\
&\iff t \in T && \text{by definition of } T
\end{aligned}$$

In particular, $\tau_{\beta_2,\tau_{\beta_1,t}}$ maps onto T. Since $\tau_{\beta_2,t}, \tau_{\beta_1,t} : \mathbb{N} \to \mathbb{N}$ are both assumed to be strictly increasing mappings, so is their composite. In particular, $\tau_{\beta_2,\tau_{\beta_1,t}}$ is an increasing enumeration of T. Since we are assuming that $\tau_{\beta,t}$ is the *unique* increasing enumeration of T, $\tau_{\beta_2,\tau_{\beta_1,t}} = \tau_{\beta,t}$, and $\pi_{\beta_1,*}\left(\pi_{\beta_2,*}\left(X_*\right)\right) = \pi_{\beta_1,*}\left(X^2_*\right) = X^2_{\tau_{\beta_1,*}} = X_{\tau_{\beta_2,\tau_{\beta_1,t}}} = X_{\tau_{\beta,t}}$. □

Theorem 5.27 allows us to easily show that the reduction construction on matrices of section 5.2 is "natural", as well.

Theorem 5.30. *If $M \in \mathrm{Mat}_n(\mathbb{R})$ is Markov, $s = s_1 \cup s_2$ is open with respect to M, (M_1, p_1, i_1) is the reduction of M with respect to s_1, (M_2, p_2, i_2) is the reduction of M_1 with respect to $s'_2 = \overline{s}_1^{-1}(s_2)$, and $\left(\widehat{M}, p, i\right)$ is the reduction of M with respect to s, then $M_2 = \widehat{M}$, $p = p_2 p_1$, and $i = i_1 i_2$.*

Proof. First, notice that, by Theorem 5.10, s'_2 is open with respect to M_1, so that the statement of the Theorem makes sense. If ι is the identity on S_n, and v is some n-dimensional distribution, we may define a chain, X_*, which is ι-consistent with M such that $\Pr[X_0 = j] = v_j, \forall j \in S_n$. By Theorem 5.27, M_1 is ι_1-consistent with $\overline{X}_* = \pi_{s_1}(X_*)$ and $\Pr\left[\overline{X}_0 = \iota_1(j)\right] = (p_1 v)_j$, where $\iota_1 = \overline{s}_1$. Likewise, M_2 is ι_2-consistent with $\widetilde{X}_* = \pi_{\iota_1\left(s'_2\right)}\left(\overline{X}_*\right) = \pi_{\overline{s}_1 \cap s_2}\left(\overline{X}_*\right)$ and $\Pr\left[\widetilde{X}_0 = \iota_2(j)\right] = (p_2 p_1 v)_j$, where $\iota_2 = \iota_1 \overline{s}'_2 = \overline{s}$. Thus, if we let $\widehat{X}_* = \pi_s(X_*)$, by Theorem 5.29, $\widetilde{X}_* = \pi_{\overline{s}_1 \cap s_2}\left(\pi_{s_1}(X_*)\right) = \pi_s(X_*) = \widehat{X}_*$ with probability 1. Since $\widetilde{X}_*$ is ι_2-consistent with M_2, and $\widehat{X}_*$ is ι_2-consistent with $\widehat{M}$, $\widehat{M} = M_2$.

For any $k \in S_n$, if $v = e_k$, then $(p_2 p_1 e_k)_j = \Pr\left[\widetilde{X}_0 = \overline{s}(j)\right] = \Pr\left[\widehat{X}_0 = \overline{s}(j)\right] = (pe_k)_j$, so that $p = p_2 p_1$. Finally, by Theorem 5.12, $i(v)$ is the unique extension of an eigenvector $v \in \ker\left(\widehat{M} - I\right)$ to an eigenvector in $\ker(M - I)$. Likewise, $i_1 i_2(v)$ is an extension of an eigenvector $v \in \ker(M_2 - I)$ first to an eigenvector in $\ker(M_1 - I)$, and then to an eigenvector in $\ker(M - I)$. Therefore, since $\widehat{M} = M_2$, by uniqueness, $i_1 i_2(v) = i(v)$. □

Part II

Computing Stochastically Stable Distributions

Chapter 6

Markov Chain Tree Theorem

In this chapter, we sharpen a result, often known as the Markov Chain Tree Theorem, proven for example by Freidlin and Wentzell (Friedlin and Wentzell, 1984), specifically for irreducible Markov matrices. It gives a combinatorial formula for the unique stable distribution of an irreducible Markov matrix. Because this theorem will form the basis of all key results in Chapter 7, we give a detailed proof. Moreover, because wish wish to apply it to *unichain* Markov matrices, we generalize the theorem to that setting. We present a novel proof which exploits the properties of the determinant function.

6.1 Directed Spanning Trees

As given by Theorem 6.17, the Markov Chain Tree Theorem gives a combinatorial formula for the unique stable distribution of a unichain Markov matrix in terms of the weights of its directed spanning subtrees. In this section, we will:

- define what we mean by a directed tree and show how they are intimately related with unichain Markov matrices,
- show how we may enumerate all directed trees on n vertices by certain class of functions on S_n, and
- define a vector, w_M, for any Markov matrix, M, in terms of the collection of all directed spanning subtrees of $G_-(M)$, which will turn out to be proportional to the stable distribution of M.

6.1.1 DST Facts

A directed graph G that contains a unique directed walk from any vertex in G to some distinguished vertex v has been called an "oriented" tree (Knuth, 1997, p. 373). We will refer to such a graph as a *directed tree*. We will also describe it as being *rooted at* v. This terminology is justified by the following theorem:

Theorem 6.1. *If $G = (V, E)$ is a directed tree rooted at v, then*

- *there is a well-defined function $l_G : V \to \mathbb{N}$ such that $l_G(v) = 0$ and for all $(u, w) \in E$, $l_G(u) = l_G(w) + 1$;*
- *v has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge; and*
- *the undirected graph associated with G cannot contain any cycles, i.e., it is a tree.*

Proof. Define $l_G : V \to \mathbb{N}$ such that $l_G(u)$ is the length of the unique walk in G from u to v. By definition, $l_G(v) = 0$. Given a walk between two vertices, say, from u to w, the walk from u to v must be the concatenation of the given walk (from u to w) and the walk from w to v. This holds for any edge (u, w), so $l_G(u) = l_G(w) + 1$.

The vertex v cannot have an outgoing edge, (v, u), since that would imply that $0 = l_G(v) = l_G(u) + 1$, and $l_G(u) = -1$, which is impossible since $l_G(u)$ is a length. Since there is a walk from every other u to v, every other u must have at least one outgoing edge. It cannot have more than one, however, because that would imply two distinct walks from u to v.

Since l_G is strictly decreasing along any walk, G cannot contain a (directed) cycle. In particular, it cannot contain any self-loops. More generally, the associated undirected graph, G_-, cannot contain an (undirected) cycle. If it did, we could find a vertex, u, in the cycle such that $l_G(u)$ is maximum among all vertices in the cycle. Since G contains no self-loops, the cycle has length at least 1, and there are two edges in the cycle incident with u in G'. These edges correspond to directed edges in G. Since $l_G(u)$ is maximum, u must be the starting vertex for both edges. But u has only one outgoing edge, so this is a contradiction. □

We know from Lemma 1.1 that every directed graph contains at least one closed class. The main result established in this section is: if a directed graph contains *exactly* one closed

class, then it contains directed spanning (i.e., containing all vertices) subtrees rooted at each of the vertices in that class.

We will say that a graph, $G = (V, E)$, contains a *star at* v iff for every $w \in V$, $(w, v) \in E$. Likewise, we will say that G is *starry* iff it contains a star at some $v \in V$. Note that G_T contains a star at v iff there is a walk from every other vertex in G to v.

Lemma 6.2. *For any directed graph $G = (V, E)$, G_T contains a star at the vertex v iff G contains exactly one closed class and v is a vertex in that class.*

Proof. Assume G contains exactly one closed class, $\mathcal{C}$, and choose an arbitrary vertex $v \in \mathcal{C}$. Now for any other vertex $w \in V$, either $w \in \mathcal{C}$ or $w \notin \mathcal{C}$. If $w \in \mathcal{C}$, w and v are in the same SCC, so there is a walk from w to v. If $w \notin \mathcal{C}$, w is transient, since G contains only one closed class, so by Lemma 1.1 there is a path from w terminating in a closed class, which must be $\mathcal{C}$. Let $u \in \mathcal{C}$ be the vertex at which this path terminates. Since u and v are in the same SCC, there is a walk from u to v, and therefore there is a walk from w to v. So there is a walk in G from every vertex to v, and G_T contains a star at v.

Conversely, assume that G_T contains a star at v. First, by Lemma 1.1, G_T must contain at least one closed class, call it $\mathcal{C}$. Second, v must be in $\mathcal{C}$. If it were not, there could be no walk from $w \in \mathcal{C}$ to v, since $\mathcal{C}$ has no outgoing edges. Third, there cannot be more than one closed class, since by the same argument v would have to be in all of them. □

Lemma 6.2 says that if G contains exactly one closed class, then it contains a directed walk from any vertex in G to each vertex in that class. In the remainder of this section, we establish a stronger result, namely that the assumption of exactly one closed class in G implies that for each vertex v in the closed class, G contains a subgraph G' in which there is a *unique* directed walk (which is necessarily a path) from any vertex in G' to v (i.e., G contains directed spanning subtrees rooted at each of the vertices in the closed class).

Lemma 6.3. *For any directed graph, G, G_T contains a star at the vertex v iff G contains a directed spanning subtree rooted at v.*

Proof. Assume that G contains a directed spanning subtree rooted at v. By definition, it then contains a (unique) directed walk from any vertex to v. This means that G_T contains an edge from any vertex to v. In other words, G_T contains a star at v.

Conversely, if G_T contains a star at v, we can use the well-known graph algorithm breadth-first search to construct a directed spanning subtree rooted at v. Breadth-first search

starts with the root node, v, in a graph, $G = (V, E)$. Each vertex reached by the algorithm is first discovered, then placed in queue, then processed when it is dequeued. To process a vertex, the algorithm discovers and enqueues all undiscovered vertices adjacent to the current vertex. It then dequeues the next vertex for processing. In this way, the algorithm processes all vertices a certain number of edges away from v before descending to the next level of depth. The algorithm uses a "color" decoration to guarantee that no vertex is processed more than once. Vertices not yet discovered are WHITE, vertices discovered but not yet processed are GRAY, and processed vertices are BLACK.

The pseudo-code below is a modification of the basic algorithm. Since the edges in a directed spanning tree point toward the root, this algorithm traverses edges backwards, that is, at any iteration it discovers a vertex u iff there is an edge pointing from u to the vertex currently being processed. In addition, it keeps track of the edges traversed this way in the set E'. We will argue that the returned graph $G' = (V, E')$, clearly a subgraph of G, is a directed spanning tree contained in G.

Algorithm 1 BFS Tree

```
Q = new Queue()
E' = ∅
for (u ∈ V)
    color[u] = WHITE
color[v] = GRAY
Q.enqueue(v)
while (!Q.isEmpty()) {
    u = Q.dequeue()
    for (; w | (w, u) ∈ E; ) {
        if (color[w] = WHITE) {
            color[w] = GRAY
            E'.insert((w, u))
            Q.enqueue(w)
        }
    }
    color[u] = BLACK
}
return(G' = (V, E'));
```

For any non-root vertex $u \neq v \in V$,

- u is eventually discovered (turned GRAY) by the algorithm. Since G_T contains a

star at v, there is a path from u to v in G. Enumerate the vertices on this path $p_1, \ldots, p_l$ with $p_1 = v$, $p_l = u$, and $(p_i, p_{i-1}) \in E$. Claim: every vertex on the path will be discovered. $p_1 = v$ is discovered in the initialization step. Suppose p_{i-1} is discovered. Then p_{i-1} is enqueued, which guarantees that it will eventually be processed. When it is processed, p_i is examined, since $(p_i, p_{i-1}) \in E$. If p_i is GRAY, it has already been discovered. If not, it is discovered, turned GRAY, and enqueued. So by induction, $p_l = u$ will eventually be discovered.

- there is a path from u to v in G'. Enumerate the non-root vertices $w_1, \ldots, w_m$ in the order they are discovered. The first, w_1, is discovered because there is an edge, and therefore a path, (w_1, v) in G, and this edge is added to E'. Suppose there is a path from w_k to v in G' for all $k < i$. Then when w_i is discovered, an edge (w_i, w_j) is added to E' from w_i to the vertex currently being processed. Since all vertices are discovered before they are processed, $j < i$. So there is a path from w_j to v, and with the new edge, there is a path from w_i to v in G'. By induction, there is a path from u to v, since u is discovered.

- this path is unique. When a non-root vertex is discovered, one edge is added to E' out of that vertex. Since each vertex is discovered exactly once, exactly one edge leaves each non-root vertex in G'. The path from u to v uses the only edge out of each vertex in the path, so it must be unique.

Thus, G' is a directed spanning tree by definition.

□

Lemma 6.2 says that G contains a star at v iff G contains exactly one closed class and v is in that class. Lemma 6.3 says that G_T contains a star at v iff G contains a directed spanning subtree rooted at v. Therefore,

Theorem 6.4. *A directed graph G contains a directed spanning subtree rooted at a vertex v iff G contains exactly one closed class, and v is a vertex in that class.*

6.1.2 DST Construction

We now give a way to represent the set of directed spanning trees of the complete graph on n vertices in terms of certain mappings. This will allow us to give a constructive proof, in

Section 6.2, of the Markov Chain Tree Theorem using determinants.

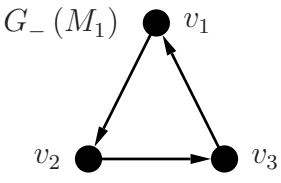
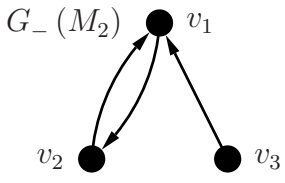
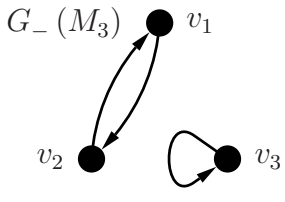
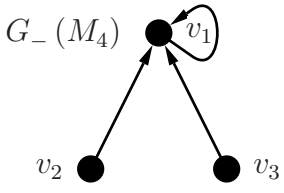
Among all directed unweighted graphs on n vertices, we restrict attention to those which are *1-regular*, that is, those in which each vertex has a unique outgoing edge. There is a 1-1 correspondence between such graphs and the set T of mappings $\sigma : S_n \to S_n$, which we will suggestive refer to as "1-regular" mappings. Specifically, for a 1-regular graph $G = (V, E)$, let $\mathrm{map}(G) = \sigma$ such that $\sigma(i) = j$ iff $(v_i, v_j) \in E$. Conversely, any such σ defines a 1-regular graph, $G_-(\sigma)$, such that $(v_i, v_j) \in E$ iff $\sigma(i) = j$. Clearly, $G_-(\mathrm{map}(G)) = G$ and $\mathrm{map}(G_-(\sigma)) = \sigma$.

There is also a 1-1 correspondence between 1-regular mappings and the set $\mathcal{M}$ of $n \times n$ square Markov matrices with a single non-zero entry (i.e., 1) in each column, which we will again refer to as "1-regular". To any 1-regular mapping $\sigma \in T$, we associate a 1-regular matrix $\mathrm{mat}(\sigma) \in \mathcal{M}$ as follows: $(\mathrm{mat}(\sigma))_{i,j} = 1$ iff $\sigma(j) = i$. Observe that each column j of $\mathrm{mat}(\sigma)$ is the standard basis vector $e_{\sigma(j)}$, so $\mathrm{mat}(\sigma) = \left(e_{\sigma(1)} \ldots e_{\sigma(n)}\right)$. Conversely, for any matrix $M \in \mathcal{M}$ we can define $\mathrm{map}(M) \in T$ such that $(\mathrm{map}(M))(j) = i$ iff $M_{i,j} = 1$. Again, mat and map are inverses and so give a 1-1 correspondence between the set of 1-regular mappings and $\mathcal{M}$.

Finally, note that these correspondences induce a 1-1 correspondence between the set of 1-regular matrices and 1-regular graphs, which is just the usual procedure of associating with a matrix M its unweighted graph $G_-(M)$. We will also define $\mathrm{mat}(G)$ for any 1-regular graph G to be the corresponding 1-regular matrix.

Four such 1-regular matrices, with their corresponding graphs, are shown in Table 6.1.

Table 6.1: Four 1-Regular Markov Matrices and Graphs. $M_3 \in \mathcal{M}_3$ (equivalently, $\mathrm{map}(M_3) \in T_3$) and $M_4 \in \mathcal{M}_1$ (equivalently, $\mathrm{map}(M_4) \in T_1$).

$M_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$M_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$M_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$M_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
			

For $i \in S_n$, define $\mathcal{M}_i = \{M \in \mathcal{M} \mid M_{j,j} = 1 \text{ iff } j = i\}$ and $T_i = \{\text{map}(M) \mid M \in \mathcal{M}_i\}$. These sets correspond to 1-regular graphs with exactly one self-loop at v_i, such as M_3 and M_4 in Table 6.1. Now restrict attention further to graphs of unichain Markov matrices, defining $\overline{\mathcal{M}}_i = \{M \in \mathcal{M}_i \mid M \text{ unichain}\}$ and $\overline{T}_i = \{\text{map}(M) \mid M \in \overline{\mathcal{M}}_i\}$. By definition, for every $M \in \overline{\mathcal{M}}_i$, $G_-(M)$ has exactly one closed class, which must be the singleton, $\{v_i\}$. For example, $M_4 \in \overline{\mathcal{M}}_1$. Notice that while M_2 is unichain, it does not contain a self-loop. Likewise, while M_3 has a self-loop at v_3, it has two closed classes, $\{v_1, v_2\}$ and $\{v_3\}$, so it is not unichain.

Table 6.2 depicts all members of the set $\mathcal{M}_1$ with vertices in S_3. Note that the diagonal entries and the entries in the first column of each matrix are determined by the definition of $\mathcal{M}_1$. This leaves two possibilities each for the two non-zero entries in columns 2 and 3, resulting in the four members. Of these four, M_1, M_2, and M_3 are unichain (each has only one closed class, $\{v_1\}$) so they are members of $\overline{\mathcal{M}}_1$. M_4, on the other hand, has two closed classes, $\{v_1\}$ and $\{v_2, v_3\}$, so it is not a member of $\overline{\mathcal{M}}_1$.

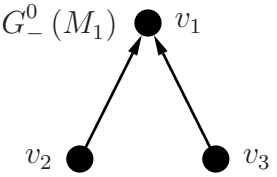
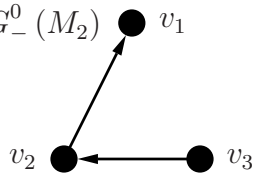
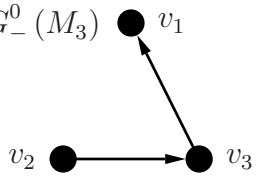
Table 6.2: The set $\mathcal{M}_1$ of 3×3 matrices. $M_1, M_2, M_3 \in \overline{\mathcal{M}}_1$ (equivalently, $\text{map}(M_1), \text{map}(M_2), \text{map}(M_3) \in \overline{T}_1$).

$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$M_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$M_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$M_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
$G_-(M_1)$: v_1, v_2, v_3	$G_-(M_2)$: v_1, v_2, v_3	$G_-(M_3)$: v_1, v_2, v_3	$G_-(M_4)$: v_1, v_2, v_3

Now for $\sigma \in \overline{T}_i$, let $G^0_-(\sigma)$ be the graph obtained by removing the self-loop at v_i from $G_-(\sigma)$. We will call the set of all such graphs $D_i = \{G^0_-(\sigma) \mid \sigma \in \overline{T}_i\}$. Notice that these graphs are directed spanning trees. In fact, this method constructs all directed spanning trees rooted at v_i, as shown in the following theorem. An illustration is given in Table 6.3, showing the associated Markov matrices, $\text{mat}(\sigma) \in \overline{\mathcal{M}}_i$ for $\sigma \in \overline{T}_i$.

Theorem 6.5. *D_i is the set of all directed spanning trees on n vertices rooted at v_i.*

Table 6.3: The set D_1 of directed spanning trees with vertices in S_3 rooted at v_1 and their associated matrices. These are the elements of $\overline{\mathcal{M}}_1$ with the self-loops at v_1 removed.

$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$M_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$M_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
$G^0_-(M_1)$ 	$G^0_-(M_2)$ 	$G^0_-(M_3)$ 

Proof. For any mapping $\sigma \in \overline{T}_i$, $\mathrm{mat}(\sigma)$ is unichain, so $G_-(\mathrm{mat}(\sigma)) = G_-(\sigma)$ contains exactly one closed class, which must be the vertex v_i with the self-loop. By Theorem 6.4, $G_-(\sigma)$ contains a directed spanning tree subgraph rooted at v_i. By removing the self-loop at v_i, we obtain $G^0_-(\sigma)$, which, since $G_-(\sigma)$ is 1-regular, has no outgoing edges from v_i and one outgoing edge from every other vertex. But by Theorem 6.1, the directed spanning subtree has no outgoing edges from the root, v_i, and one outgoing edge from every other vertex. $G^0_-(\sigma)$ is the only possible subgraph with these properties, so it is a directed spanning tree.

Moreover, every such tree can be constructed in this fashion. Given a directed spanning tree G' rooted at v_i, add a self-loop at v_i to obtain a 1-regular graph, G, with associated σ. Since the directed spanning tree G' is a subgraph of G, by Theorem 6.4, G contains exactly one closed class. In particular, $\mathrm{mat}(\sigma)$ is unichain, i.e., $\mathrm{mat}(\sigma) \in \overline{\mathcal{M}}_i$, so that $\sigma \in \overline{T}_i$ and $G^0_-(\sigma) = G'$. Thus, $G' \in D_i$, and D_i is the set of all directed spanning trees rooted at v_i. □

6.1.3 The vector w_M

Now that we have a construction for the set of directed spanning trees, for any given Markov matrix, M, we may enumerate the directed spanning subtrees of $G_-(M)$, and their associated "weights". We will see that when M is unichain, these are closely related to the stable distribution of M.

Given any Markov matrix, M, and any $\sigma \in T_i$, we define

$$W(M, \sigma) = \Pi_{j \neq i} M_{\sigma(j),j} \tag{6.1}$$

Intuitively, this is the "total" weight in $G(M)$ of the edges in $G^0_-(\sigma)$ (where we aggregate by multiplication rather than addition), which is a directed spanning tree when $\sigma \in \overline{T}_i$. Notice how we take the product over $j \neq i$, so that the "total" weight excludes the weight on the self-loop at i in $G_-(\sigma)$. Notice further that, while we will usually apply this definition to $\sigma \in \overline{T}_i$, $W(M, \sigma)$ is well-defined for any $\sigma \in T_i$.

Given a Markov matrix M, we now define the vector w_M such that

$$(w_M)_i = \sum_{\sigma \in \overline{T}_i} W(M, \sigma) \tag{6.2}$$

The ith entry of w_M is the sum of the "total" weights in $G(M)$ of all directed spanning subtrees rooted at v_i.

Example 6.6. Throughout the remainder of this chapter, we will use the Markov matrix $M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ as a running example. To calculate $(w_M)_1$, we must sum over all $\sigma \in \overline{T}_1$, which correspond to the matrices $M_1, M_2, M_3 \in \overline{\mathcal{M}}_1$ enumerated in Table 6.2.

First, calculate $W(M, \sigma_1)$, for $\sigma_1 = \text{map}(M_1)$; in particular, $\sigma_1(1) = 1$, $\sigma_1(2) = 1$, and $\sigma_1(3) = 1$. Here, $W(M, \sigma_1) = \Pi_{j \neq 1} M_{\sigma_1(j),j} = M_{\sigma_1(2),2} M_{\sigma_1(3),3} = M_{1,2} M_{1,3} = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$. Similarly, $W(M, \sigma_2) = \Pi_{j \neq 1} M_{\sigma_2(j),j} = M_{\sigma_2(2),2} M_{\sigma_2(3),3} = M_{1,2} M_{2,3} = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$, and $W(M, \sigma_3) = \Pi_{j \neq 1} M_{\sigma_3(j),j} = M_{\sigma_3(2),2} M_{\sigma_3(3),3} = M_{3,2} M_{1,3} = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$.

Finally, $(w_M)_1 = \sum_{\sigma \in \overline{T}_1} W(M, \sigma) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$. Repeating this process for $\sigma \in \overline{T}_2$ and $\sigma \in \overline{T}_3$, we find that $w_M = \begin{pmatrix} \frac{3}{8} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$. □

Lemma 6.7. *For any $n \times n$ Markov matrix M, all $i \in S_n$, and $\sigma \in \overline{T}_i$, $W(M, \sigma) \neq 0$ iff the directed spanning tree associated to σ, $G^0_-(\sigma)$, is a subgraph of $G_-(M)$.*

Proof. Abbreviate the directed spanning tree associated to σ by G. We may enumerate all the edges of G as $\left(v_j, v_{\sigma(j)}\right)$ for $j \in S_n \setminus \{i\}$. An edge, $\left(v_j, v_{\sigma(j)}\right)$, is in $G_-(M)$ iff it is in $G(M)$ iff it has positive weight, i.e., $M_{\sigma(j),j} > 0$. Thus, G is a subgraph of $G_-(M)$ iff $M_{\sigma(j),j} > 0$ for all $j \in S_n \setminus \{i\}$ iff $0 \neq \Pi_{j \neq i} M_{\sigma(j),j} = W(M, \sigma)$. □

By Lemma 6.7, when $G^0_-(\sigma)$ is not a subgraph of $G_-(M)$, the corresponding term in $\sum_{\sigma\in\overline{T}_i} W(M,\sigma)$ is zero. For M Markov, define $T(M,i)$ to be only those mappings σ in $\overline{T}_i$ whose associated direct spanning trees are subgraphs of $G_-(M)$: i.e.,

$$T(M,i) = \{\sigma \in \overline{T}_i \mid G^0_-(\sigma) \subset G_-(M)\} \tag{6.3}$$

We may now give an equivalent definition of w_M with zero terms removed from the sum:

$$(w_M)_i = \sum_{\sigma\in T(M,i)} W(M,\sigma) \tag{6.4}$$

We will use this definition from now on.

Theorem 6.8. *The vector $w_M \neq 0$ iff M is unichain. Specifically, if M is unichain, $(w_M)_i \neq 0$ iff v_i is in the closed class of $G(M)$.*

Proof. Suppose M is unichain. By Theorem 6.4 there exists a subgraph $G'_i \subset G_-(M)$ which is a directed spanning tree rooted at the vertex v_i iff v_i is in the closed class of $G_-(M)$. If v_i is in the closed class, let $\sigma_i \in \overline{T}_i$ be mapping associated to G'_i (so that $G'_i = G^0_-(\sigma_i)$). Then $\sigma_i \in T(M,i)$, and there is at least one term in the sum $\sum_{\sigma\in T(M,i)} W(M,\sigma)$. Since this is a sum of positive terms, $(w_M)_i \neq 0$. If v_i is not in the closed class, $G(M)$ has no directed spanning subtree rooted at v_i, so $T(M,i)$ is empty and $(w_M)_i = 0$.

If M is not unichain, by Theorem 6.4 $G(M)$ has no directed spanning subtree rooted at any vertex. So $T(M,i)$ is empty for all i, and $w_M = 0$. □

6.2 A Proof Using Determinants

Having established all necessary combinatorial definitions in section 6.1, we now move on the proof of the Markov Chain Tree Theorem. The proof will depend primarily on the multi-linearity of the determinant function from linear algebra. Thus, we will begin by reviewing basic facts and definitions associated with the determinant function. Specifically, we will:

- review the basic properties of the determinant,
- define what we mean by minors, cofactors, and the adjoint of a matrix,

- show how the vector, w_m, from section 6.1 naturally occurs as the diagonal of the adjoint of the laplacian of a unichain Markov matrix, M, and
- use linear algebra to show that this must then be proportional to the stable distribution of M.

6.2.1 Determinants

We begin by recalling some basic facts regarding the determinant function on square matrices. For notational convenience, we will sometimes write $v_1 \wedge \cdots \wedge v_n$ for the determinant of the $n \times n$ square matrix with v_is as columns, where $v_i \in \mathbb{R}^n$.

The determinant of a 2×2 matrix, $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$. The determinant of an $n \times n$ matrix, N, for $n > 2$ can be calculated recursively as follows, using the Laplace expansion formula. The $(i,j)^{\text{th}}$ *minor* of N, $N^{i,j}$, is the $(n-1) \times (n-1)$ matrix obtained by removing the i^{th} row and j^{th} column from N. The $(i,j)^{\text{th}}$ cofactor of N, $C_N^{i,j} = (-1)^{i+j}|N^{i,j}|$. Now, for any row i or column j of N,

$$|N| = \sum_{k=1}^{n} N_{i,k} C_N^{i,k} = \sum_{k=1}^{n} N_{k,j} C_N^{k,j} \,. \tag{6.5}$$

Example 6.9. For example, the determinant of the matrix $M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ can be calculated by applying the Laplace expansion along the first column: $|M| = M_{1,1}C_M^{1,1} + M_{2,1}C_M^{2,1} + M_{3,1}C_M^{3,1} = 0 \times C_M^{1,1} + 1 \times (-1)^{2+1} \begin{vmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} + 0 \times C_M^{3,1} = -1(\frac{1}{2} \times \frac{1}{2} - \frac{1}{4} \times \frac{1}{2}) = -\frac{1}{8}$. □

Theorem 6.10. *The determinant function on square matrices has the following well-known properties (see, for example, (Horn and Johnson, 1985)):*

a) $|NM| = |N||M|$, *or equivalently, setting* $v_i = Me_i$, *so that* $|M| = v_1 \wedge \cdots \wedge v_n$,

$$(Nv_1) \wedge \cdots \wedge (Nv_n) = |N|\, v_1 \wedge \cdots \wedge v_n \tag{6.6}$$

b) $|I| = 1$ *and* $\left|M^{-1}\right| = |M|^{-1}$.

c) $|N| = 0$ *iff* N *is not invertible.*

d) $|M^{-1}NM| = |N|$, *and in particular, for any permutation matrix* P, $\left|P^t MP\right| = |M|$.

e) $\left|M^t\right| = |M|$

f) *If* N *is block-triangular with square diagonal blocks,* N_i, *then* $|N| = \prod_i |N_i|$.

g) *The determinant function is* multi-linear, *meaning it is linear in each "factor", i.e.,*

$$v_1 \wedge \cdots \wedge (\alpha v_i + w) \wedge \cdots \wedge v_n = \alpha \left(v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_n\right) + v_1 \wedge \cdots \wedge w \wedge \cdots \wedge v_n$$

Another key property of the determinant involves the adjoint operator, adj. The *adjoint* of a matrix is the transpose of its matrix of cofactors, i.e., for any $n \times n$ square matrix, N,

$$(\mathrm{adj}\, N)_{i,j} = C_N^{j,i} = (-1)^{i+j} \left|N^{j,i}\right| \tag{6.7}$$

The adjoint satisfies the following equations (Wicks, 1996):

$$\mathrm{adj}(N)\, N = |N|\, I = N\, \mathrm{adj}(N) \tag{6.8}$$

These equations are equivalent to the Laplace expansion formula for determinants.

Example 6.11. Recall our matrix $M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

$M^{1,1} = \begin{pmatrix} 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, and $\mathrm{adj}(M)_{1,1} = (-1)^{1+1} \left|M^{1,1}\right| = -\frac{1}{8}$.

Similarly, $M^{1,2} = \begin{pmatrix} 1 & \frac{1}{4} \\ 0 & \frac{1}{2} \end{pmatrix}$, and $\mathrm{adj}(M)_{2,1} = (-1)^{1+2} \left|M^{1,2}\right| = -\frac{1}{2}$.

Continuing this process, we find that $\mathrm{adj}(M) = \begin{pmatrix} -\frac{1}{8} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}$.

Now it is easy to check that $\mathrm{adj}(M)M = M\,\mathrm{adj}(M) = \begin{pmatrix} -\frac{1}{8} & 0 & 0 \\ 0 & -\frac{1}{8} & 0 \\ 0 & 0 & -\frac{1}{8} \end{pmatrix} = |M|\, I$. □

6.2.2 The Stable Distribution

Define $\overline{w}_M$ to be the vector consisting of the diagonal entries of $\mathrm{adj}\,(\Lambda)$. That is, $(\overline{w}_M)_i \equiv \mathrm{adj}(\Lambda)_{i,i} = C_\Lambda^{i,i} = \left|\Lambda^{i,i}\right|$. In this section, we will show that $\overline{w}_M$ is closely related to the vector w_M defined earlier. This will lead to a formula for the stable distribution of a unichain Markov matrix in terms of its directed spanning trees.

In order to proceed, we need a bit of additional notation, let $R_i\,(N)$ denote the result of replacing the i^{th} column of N by the standard basis vector, e_i.[1]

Example 6.12. For M in the examples above,

$\Lambda = M - I = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{4} \\ 1 & -1 & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$, and $\mathrm{adj}\,(\Lambda) = \begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, so $\overline{w}_M = \begin{pmatrix} \frac{3}{8} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$.

Further, $R_1\,(\Lambda) = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & -1 & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$. Computing the determinant of this matrix (using the Laplace expansion formula) yields $|R_1\,(\Lambda)| = 1 \times \begin{vmatrix} -1 & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -1 \times -\frac{1}{2} - \frac{1}{4} \times \frac{1}{2} = \frac{3}{8} = (\overline{w}_M)_1$, which holds in general, as the following lemma shows. □

Lemma 6.13. *For any Markov matrix, M,* $(\overline{w}_M)_i = |R_i\,(\Lambda)|$.

Proof. By the Laplace expansion formula for determinant along the i^{th} column, $|R_i\,(\Lambda)| = \sum_{k=1}^n (R_i(\Lambda)_{k,i}) C_{R_i(\Lambda)}^{k,i} = C_{R_i(\Lambda)}^{i,i} = (-1)^{2i} |R_i(\Lambda)^{i,i}| = (-1)^{2i} |\Lambda^{i,i}| = \mathrm{adj}(\Lambda)_{i,i} = (\overline{w}_M)_i$. □

From our examples, it appears that $\overline{w}_M = w_M$. We will now work to show that this is (almost) true, in general. We will do this by gradually rewriting $\overline{w}_M$ via a series of lemmas, until we obtain w_M.

[1] More formally, $R_i\,(N) = N + (I - N)e_i e_i^t$, so that $R_i\,(N)\,e_j = Ne_j + (I - N)e_i e_i^t e_j = Ne_j + (I - N)e_i[i = j] = e_I[i = j] + N(e_j - e_i[i = j]) = e_I[i = j] + Ne_j[i \neq j]$.

Lemma 6.14. *For any Markov matrix, M,* $(\overline{w}_M)_i = \sum_{\sigma\in T_i} W(M,\sigma)\,|R_i(\Lambda(\mathrm{mat}(\sigma)))|$, *where, by our usual convention,* $\Lambda(\mathrm{mat}(\sigma)) = \mathrm{mat}(\sigma) - I$.

Proof. By Lemma 6.13, remembering that $\Lambda \equiv M - I$, we may first write $(\overline{w}_M)_i$ in terms of the columns of $R_i(\Lambda)$,

$$\begin{aligned}(\overline{w}_M)_i &= |R_i(\Lambda)| \\ &= R_i(\Lambda)\,e_1 \wedge \cdots \wedge R_i(\Lambda)\,e_i \wedge \cdots \wedge R_i(\Lambda)\,e_n \\ &= \Lambda e_1 \wedge \cdots \wedge e_i \wedge \cdots \wedge \Lambda e_n \end{aligned} \tag{6.9}$$

Since the columns of Λ sum to 0 (i.e., are in $\ker J$), we may write the jth column of Λ, Λe_j, in terms of $\overline{e}_{i,j} \equiv e_i - e_j$ for $i \neq j$ (i.e., a choice of basis for $\ker J$). We begin, as follows:

$$\Lambda e_j = M e_j - e_j = \sum_{i=1}^{n} M_{i,j} e_i - e_j = \sum_{i\neq j} M_{i,j} e_i + \left(M_{j,j} - 1\right) e_j\,.$$

Since the jth column of M sums to 1, we obtain the desired expansion:

$$\Lambda e_j = \sum_{i\neq j} M_{i,j} e_i + \left(-\sum_{i\neq j} M_{i,j}\right) e_j = \sum_{i\neq j} M_{i,j}(e_i - e_j) = \sum_{i\neq j} M_{i,j}\overline{e}_{i,j} \tag{6.10}$$

Applying Equation 6.10 to Equation 6.9, we have, by the multi-linearity of the determinant,

$$\begin{aligned}(\overline{w}_M)_i &= \sum_{s_1\neq 1} M_{s_1,1}\overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \sum_{s_n\neq n} M_{s_n,n}\overline{e}_{s_n,n} \\ &= \sum_{s_1\neq 1} \cdots \sum_{s_n\neq n} \left(M_{s_1,1}\cdots M_{s_n,n}\right)\left(\overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{s_n,n}\right) \\ &= \sum_{s_1\neq 1} \cdots \sum_{s_n\neq n} \left(\Pi_{j\neq i} M_{s_j,j}\right)\left(\overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{s_n,n}\right) \end{aligned} \tag{6.11}$$

We now apply the substitution $s_j = \sigma(j)$, so that each choice of values for the summation variables, $\{s_1, \ldots, \widehat{s_i}, \ldots, s_n\}$, represents a unique choice of $\sigma : S_n \setminus \{i\} \to S_n$. No choice of $\sigma(i)$ is made because Equation 6.11 does not include a summation over s_i. Notice that the sum now requires $\sigma(j) \neq j$ for all $j \neq i$. We may also require $s_i = \sigma(i) = i$ to obtain a unique choice of $\sigma : S_n \to S_n$, such that $\sigma(j) = j$ iff $j = i$, i.e., $\sigma \in T_i$. Therefore, Equation 6.11 may be rewritten as

$$\begin{aligned}(\overline{w}_M)_i &= \sum_{\sigma\in T_i} \left(\Pi_{j\neq i} M_{\sigma(j),j}\right)\left(\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n}\right) \\ &= \sum_{\sigma\in T_i} W(M,\sigma)\left(\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n}\right) \end{aligned} \tag{6.12}$$

Now consider $\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n}$. Converting back to standard determinant notation, we have

$$\begin{aligned}
\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n} &= \left| \begin{array}{ccccc} \overline{e}_{\sigma(1),1} & \cdots & e_i & \cdots & \overline{e}_{\sigma(n),n} \end{array} \right| \\
&= \left| R_i \left(\begin{array}{ccccc} \overline{e}_{\sigma(1),1} & \cdots & \overline{e}_{\sigma(i),i} & \cdots & \overline{e}_{\sigma(n),n} \end{array} \right) \right| \\
&= \left| R_i \left(\begin{array}{ccc} e_{\sigma(1)} - e_1 & \cdots & e_{\sigma(n)} - e_n \end{array} \right) \right| \\
&= \left| R_i \left(\left(\begin{array}{ccc} e_{\sigma(1)} & \cdots & e_{\sigma(n)} \end{array} \right) - I \right) \right| \\
&= |R_i(\Lambda(\mathrm{mat}(\sigma)))| \qquad (6.13)
\end{aligned}$$

Combining Equations 6.12 and 6.13 give our desired equation

$$(\overline{w}_M)_i = \sum_{\sigma \in T_i} W(M, \sigma) \, |R_i(\Lambda(\mathrm{mat}(\sigma)))|$$

□

The formula from Lemma 6.14 may be simplified significantly, once we prove the following lemma.

Lemma 6.15. *For any $\sigma \in T_i$, $|R_i(\Lambda(\mathrm{mat}(\sigma)))| = (-1)^{n-1}$, whenever $\sigma \in \overline{T}_i$, and 0 otherwise.*

Proof. Suppose that $\sigma \in \overline{T}_i$. Consider the associated directed spanning tree, $D = G^0_-(\sigma)$. We can assign each vertex in $G_-(M)$ a number according to the length function l_G on D, given in Theorem 6.1. By sorting the vertices from low to high by their value under the length function, and renumbering the vertices in this sorted order, we achieve the property that the edge out of each non-root vertex ends at a lower numbered vertex. This is because for $(u, w) \in E$, with $u \neq v_i$, $l_G(u) = l_G(w) + 1$. Since $l_G(v_i) = 0$, v_i is renumbered as the first vertex.

From a matrix perspective, if we permute the rows and columns according to this renumbering of the vertices, the result is upper-triangular, since edges always go from a higher (column) index to a lower (row) index. Moreover, the diagonal contains all 0's except in the $(1, 1)$-entry, corresponding to the fact that the graph has a self-loop only at the root, which gets renumbered with index 1. Therefore, the same permutation of Λ is upper-triangular with -1's on the diagonal, except for a 0 in the $(1, 1)$-entry, and the same

permutation of $R_i(\Lambda)$ is upper-triangular with -1's on the diagonal, except for a 1 in the (n,n)-entry. By Theorem 6.10 d), permuting the rows and columns does not affect the determinant. So by Theorem 6.10 f), the determinant is the product of these 1×1 diagonal blocks, and $|R_i(\Lambda(\mathrm{mat}(\sigma)))| = (-1)^{n-1}$.

Now suppose that $\sigma \notin \overline{T}_i$. $M = \mathrm{mat}(\sigma)$ is not unichain, so it has at least two closed classes. Pick two such closed classes, s_1 and s_2. By Lemma 2.10, there exists a permutation matrix such that $P_s^t M P_s = \begin{pmatrix} *_0 & 0 & 0 \\ *_1 & M_1 & 0 \\ *_2 & 0 & M_2 \end{pmatrix}$, where M_1 and M_2 are the Markov principal sub-matrices corresponding to s_1 and s_2, and the $*$'s are unknown entries. In particular, it is block lower-triangular, as is $P^t \Lambda P = \begin{pmatrix} *_0 - I & 0 & 0 \\ *_1 & \Lambda_1 & 0 \\ *_2 & 0 & \Lambda_2 \end{pmatrix}$, with diagonal blocks, $D_1 = *_0 - I$, $D_2 = M_1 - I = \Lambda_1$, and $D_3 = M_2 - I = \Lambda_2$. $P^t R_i(\Lambda) P$ is also block lower-triangular with exactly one of the diagonal blocks, D_j, replaced by $R_{i'}(D_j)$, for some i' (determined by P and i) and $j = 1, \ldots, 3$.

By Theorem 6.10 d), $|R_i(\Lambda)| = \left|P^t R_i(\Lambda) P\right|$, which, by Theorem f), is either $|R_{i'}(*_0 - I)|\,|\Lambda_1|\,|\Lambda_2|$, $|*_0|\,|R_{i'}(\Lambda_1)|\,|\Lambda_2|$, or $|*_0|\,|\Lambda_1|\,|R_{i'}(\Lambda_2)|$. Since Λ_1 and Λ_2 correspond to Markov matrices M_1 and M_2, neither is invertible. In other words, both have determinant zero, so there is at least one zero term in each product, and $|R_i(\Lambda)| = 0$. □

We can now show that $\overline{w}_M$ and w_M are equal, up to a change in sign.

Theorem 6.16. *For any Markov matrix, M, $\overline{w}_M = (-1)^{n-1} w_M$.*

Proof. Focusing attention on the i^{th} components, we must show that

$$(\overline{w}_M)_i = (-1)^{n-1}(w_M)_i = (-1)^{n-1} \sum_{\sigma \in T(M,i)} W(M,\sigma)\,.$$

By Lemma 6.14,

$$(\overline{w}_M)_i = \sum_{\sigma \in T_i} W(M,\sigma)\,|R_i(\Lambda(\mathrm{mat}(\sigma)))|\,.$$

By Lemma 6.15, $|R_i(\Lambda(\mathrm{mat}(\sigma)))| = 0$, if $\sigma \notin \overline{T}_i$. Thus, this simplifies to

$$\sum_{\sigma \in \overline{T}_i} W(M,\sigma)\,|R_i(\Lambda(\mathrm{mat}(\sigma)))|\,.$$

Moreover, when $\sigma \in \overline{T}_i$, Lemma 6.15 says that $|R_i(\Lambda(\mathrm{mat}(\sigma)))| = (-1)^{n-1}$, so that this simplifies to

$$\sum_{\sigma \in \overline{T}_i} W(M,\sigma)(-1)^{n-1} = (-1)^{n-1} \sum_{\sigma \in \overline{T}_i} W(M,\sigma) = (-1)^{n-1} \sum_{\sigma \in T(M,i)} W(M,\sigma)\,,$$

as desired. □

Theorem 6.17 (Markov Chain Tree Theorem)**.** *For any unichain Markov matrix,* M*,* $\mathrm{stab}\, M = \{v_M\}$*, where*

$$(v_M)_i = \frac{1}{K}\left[\sum_{\sigma \in T(M,i)} W(M,\sigma)\right] = \frac{1}{K}(w_M)_i \tag{6.14}$$

with normalizing constant, $K = \|w_M\|_1 = \sum_{i=1}^n (w_M)_i = \sum_{i=1}^n \sum_{\sigma \in T(M,i)} W(M,\sigma)$.

Proof. By Theorem 5.14, since M has 1 closed class, $\dim \ker \Lambda$ = 1. This means that $|\Lambda| = 0$, and, by Equation 6.8, $0 = |\Lambda|\, I = \mathrm{adj}\,(\Lambda)\,\Lambda$. In other words, all rows of $\mathrm{adj}(\Lambda)$ are in $\ker \Lambda^t$. By Theorem A.1, $\dim \ker \Lambda^t = \dim \ker \Lambda = 1$, since Λ is square, and $J \in \ker \Lambda^t$, so each row of $\mathrm{adj}(\Lambda)$ must be a multiple of J, i.e., for each row, all entries in that row must be equal. This means that the columns of $\mathrm{adj}(\Lambda)$ are all identical.

Similarly, $0 = |\Lambda|\, I = \Lambda\, \mathrm{adj}(\Lambda)$. In other words, all columns of $\mathrm{adj}(\Lambda)$ are in $\ker \Lambda$. Now $\overline{w}_M$ is defined as the diagonal entries of $\mathrm{adj}\,(\Lambda)$. Since the columns of $\mathrm{adj}(\Lambda)$ are identical, $\overline{w}_M$ is also equal to each column. In particular, $\overline{w}_M \in \ker \Lambda$, and also $w_M \in \ker \Lambda$. Since M is unichain, by Theorem 6.8, $w_M \neq 0$, and hence $\{w_M\}$ is a basis for $\ker \Lambda$.

The stable distributions of M are the positive norm-1 vectors in $\ker \Lambda$. We know that w_M is positive, since its entries are the sums of products of positive weights. By letting $K = \|w_M\|_1$ and $v_M = \frac{1}{K} w_M$, we see that $v_M > 0$, and $\|v_M\|_1 = \sum_{i=1}^n (w_M)_i / K = \frac{1}{K}\|w_M\|_1 = 1$, so v_M is a stable distribution of M. Since $\dim \ker \Lambda = 1$, v_M is the unique stable distribution of M. □

Example 6.18. Completing the ongoing example, we calculate the normalization constant as follows: $K = (w_M)_1 + (w_M)_2 + (w_M)_3 = \frac{11}{8}$, so $v_M = \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix}$. Indeed, v_M is a stable

distribution, since $Mv_M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix} = \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix} = v_M$. □

Notice that Theorem 6.17 suggests that $(v_M)_i$ may be viewed as the conditional probability that a randomly selected directed spanning subtree is rooted at i, where the relative probability of each tree is given by product of the weights of its edges. This implies that, if we could efficiently sample from the corresponding distribution of directed spanning subtrees, we would have a Monte Carlo algorithm for computing v_M. While Broder (1989) provides such a sampling algorithm, it is not sufficiently fast to render the associated procedure for computing v_M competitive with other more direct techniques.

Chapter 7

Perturbed Markov Matrices

We now wish to generalize our our study of Markov matrices to the case when the entries are sufficiently "nice" functions of a non-negative parameter, ϵ, to so called "perturbed" Markov matrices (PMMs). If we denote such a matrix by M_ϵ, we will be interested in the stable distributions of M_ϵ as $\epsilon \to 0$. As such, we will need to combine the linear algebra and graph theory of Part I with some careful real analysis.

We will show that:

- a PMM, M_ϵ, has a well-defined stable distribution, v_ϵ, which is a "perturbed" matrix (i.e., column vector),
- $v_0 \equiv \lim_{\epsilon \to 0} v_\epsilon$ exists, the so-called "stochastically stable distribution" (SSD) of M_ϵ,
- v_0 only depends on M_ϵ up to an equivalence relation ("asymptotic" equality) defined over its entries,
- the asymptotic equivalence class of an entry is determined by two real-valued invariants, which we call the *resistance* and *cost* of the entry, respectively,
- likewise, the asymptotic equivalence class of a PMM, M_ϵ, may be specified by two real-valued matrices (i.e., its resistance, $R(M_\epsilon)$, and cost, $C(M_\epsilon)$),
- the two constructions from Chapter 5 (i.e., scaling and reduction), as well as the corresponding notions of equivalence and D-equivalence, generalize to PMMs, and
- by careful application of the Markov Chain Tree Theorem, we can guarantee that we only need invert *constant* matrices in our constructions.

By alternating these two constructions, we are able to give the first exact algorithm for computing v_0. Moreover, since the SSD only depends on the asymptotic equivalence class of M_ϵ, we may represent all perturbed matrices in the computation by the corresponding pair of resistance and cost matrices.

Because we will only be interested in functional values for "sufficiently small" non-negative values of ϵ, it will be useful to establish the following two conventions. If $Q(\epsilon)$ is a proposition containing the variable ϵ, we will write "$Q(\epsilon)$ for $\epsilon \succeq 0$" as an shorthand for "$\exists \delta > 0 \ \text{s.t.} \ Q(\epsilon)$ for $\epsilon \in [0, \delta]$". Likewise, "$Q(\epsilon)$ for $\epsilon \succ 0$" will mean "$\exists \delta > 0 \ \text{s.t.} \ Q(\epsilon)$ for $\epsilon \in (0, \delta]$". In other words, $\epsilon \succeq 0$ may be read as "for sufficiently small non-negative ϵ", while $\epsilon \succ 0$ will mean "for sufficiently small positive ϵ".

7.1 Exponentially Convergent Functions

In this section, we will establish the groundwork for our study of PMMs by defining precisely what we mean by "sufficiently nice functions of ϵ". The fundamental issue is that we need to restrict to a class of functions which:

- could serve as entries to a Markov matrix,
- have a well-defined limit as $\epsilon \to 0$, and
- is closed under basic algebraic operations.

In particular, we will want the collection of (Markov) *matrices*, M_ϵ, with such entries to be closed under standard matrix operations. Moreover, we will want $\mathrm{stab}\,(M_\epsilon)$ to correspond to a matrix with such entries, so that we may take limits. In addition, since we are primarily interested in functional values as $\epsilon \to 0$, they will not need to be defined for *all* non-negative ϵ. In particular, we will only be interested in such functions up to "asymptotic" equivalence.

Thus, to begin it is natural to require that the entries should at least be positive and continuous. In fact, we will be a bit more stringent. We will restrict attention to the collection of functions, $f(\epsilon)$, which are continuous for sufficiently small non-negative ϵ, and either positive for sufficiently small positive ϵ or zero for sufficiently small non-negative ϵ. We will denote this collection as $C^+[0, *]$, and, using our convention, we may define it as follows:

$$C^+[0, *] = \{f \text{ continuous for } \epsilon \succeq 0 \mid f(\epsilon) > 0, \forall\, \epsilon \succ 0 \text{ or } f(\epsilon) = 0, \forall\, \epsilon \succeq 0\}$$

Alternatively, if $C^0[0,\delta]$ denotes the set of real-valued, continuous functions on $[0,\delta]$, then

$$C^0[0,*] = \bigcup_{\delta>0} \left\{ f \in C^0[0,\delta] \mid f(\epsilon) > 0,\ \forall\, 0 < \epsilon \leq \delta \text{ or } f(\epsilon) = 0,\ \forall\, 0 \leq \epsilon \leq \delta \right\}$$

However, this collection is too big. Any $f \in C^0[0,\delta]$ and $\delta' \in (0,\delta)$, defines a restriction, $g \in C^0[0,\delta']$, so that $g(\epsilon) = f(\epsilon)$, for all $0 \leq \epsilon \leq \delta'$. We would clearly like to consider those as the "same" function.[1] Thus, we define the following relation on $C^+[0,*]$.

Definition 7.1. *For $f, g \in C^+[0,*]$, we will say that f is* asymptotically equal to *g and write $f \simeq g$ iff either:*

i) $g(\epsilon) = 0 = f(\epsilon)$ *for* $\epsilon \succeq 0$*, or*

ii) $g(\epsilon) > 0$ *for* $\epsilon \succ 0$ *and* $\lim_{\epsilon\to 0^+} \frac{f(\epsilon)}{g(\epsilon)} = 1$.

Notice that, if $g \in C^0[0,*]$, then either $g(\epsilon) = 0$ for $\epsilon \succeq 0$ or $g(\epsilon) > 0$ for $\epsilon \succ 0$ (but not both), so that Definition 7.1 makes sense.

We now show that this relation is, in fact, an *equivalence* relation, along with some other useful facts.

Lemma 7.2. *For $f_i, g_i \in C^0[0,*]$, $i = 1, 2$,*

a) if $f_1(\epsilon) = g_1(\epsilon) > 0$ for $\epsilon \succ 0$, then $f_1 \simeq g_1$;

b) $f_1 + f_2, f_1 f_2 \in C^0[0,*]$, that is, this collection of functions is "closed" under addition and multiplication;

c) $\simeq$ is an equivalence relation on $C^0[0,*]$;

d) if $f_i \simeq g_i$, $i = 1, 2$, then $f_1 f_2 \simeq g_1 g_2$, that is, $\simeq$ is "preserved" under multiplication.

Proof. The proof of part a) is almost immediate. Assuming that $f(\epsilon) = g(\epsilon) > 0$ for $\epsilon \succ 0$, $\lim_{\epsilon\to 0^+} \frac{f(\epsilon)}{g(\epsilon)} = \lim_{\epsilon\to 0^+} 1 = 1$, so that $f \simeq g$, by Definition 7.1 ii). Part b) is also clear, since the sum or product of continuous/positive functions is continuous/positive.

[1] Mathematically, we want to look at the "germs" of $C^+[0,\infty)$ at 0 (Warner, 1984).

To prove part c), we must show that $\simeq$ is reflexive, symmetric, and transitive. For any $f \in C^+[0,*]$, either $f(\epsilon) = 0$, for $\epsilon \succeq 0$, so that $f \simeq f$ by Definition 7.1 i), or $f(\epsilon) > 0$, so that, by part a), $f \simeq f$. Thus, $\simeq$ is reflexive.

Likewise, if $f \simeq g$, either $g(\epsilon) = 0$ for $\epsilon \succeq 0$, so we must be in case i) when $f(\epsilon) = 0$ for $\epsilon \succeq 0$, as well. In which case, Definition 7.1 i) gives $g \simeq f$, as well. Alternatively, $g(\epsilon) > 0$ for $\epsilon \succ 0$ and we must be in case ii). Since $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} = 1$, we cannot have $f(\epsilon) = 0$ for $\epsilon \succeq 0$, so that $f(\epsilon) > 0$ for $\epsilon \succ 0$. We may then say that $\lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{f(\epsilon)} = 1$, so that $g \simeq f$ by Definition 7.1 ii). That is, $\simeq$ is symmetric.

To finish part c), assume that $f \simeq g$ and $g \simeq h$. Since we have already shown $\simeq$ to be symmetric, we know that $h \simeq g$, as well. Now assume that $g(\epsilon) = 0$ for $\epsilon \succeq 0$, so we must be in case i), that is, we may conclude that $f(\epsilon) = 0$ and $h(\epsilon) = 0$ for $\epsilon \succeq 0$, as well. In particular, $f \simeq h$. Otherwise, $g(\epsilon) > 0$ for $\epsilon \succ 0$, and we must be in case ii). Therefore, $h(\epsilon) > 0$ for $\epsilon \succ 0$ and we may conclude that $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} = 1 = \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{h(\epsilon)}$. Thus,

$$1 = 1 \cdot 1 = \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{h(\epsilon)} = \lim_{\epsilon \to 0^+} \left(\frac{f(\epsilon)}{g(\epsilon)} \frac{g(\epsilon)}{h(\epsilon)} \right) = \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{h(\epsilon)}$$

and by Definition 7.1 ii), $f \simeq h$. Thus, we have proven that $\simeq$ is transitive.

Now we must prove part d). First consider the case when at least one of the f_i or g_i is identically 0. Assume, for example, that $f_1 \simeq 0$. Using the fact that $\simeq$ is an equivalence relation, we may reason as follows. By assumption, $g_1 \simeq f_1 \simeq 0$, so that $f_1(\epsilon) = 0 = g_1(\epsilon)$ for $\epsilon \succeq 0$. Then $f_1(\epsilon) f_2(\epsilon) = 0 = g_1(\epsilon) g_2(\epsilon)$ for $\epsilon \succeq 0$, and $f_1 f_2 \simeq 0 \simeq g_1 g_2$.

Otherwise, since $f_i \simeq g_i$, $\lim_{\epsilon \to 0^+} \frac{f_i(\epsilon)}{g_i(\epsilon)} = 1$. Therefore,

$$\lim_{\epsilon \to 0^+} \frac{f_1(\epsilon) f_2(\epsilon)}{g_1(\epsilon) g_2(\epsilon)} = \lim_{\epsilon \to 0^+} \frac{f_1(\epsilon)}{g_1(\epsilon)} \lim_{\epsilon \to 0^+} \frac{f_2(\epsilon)}{g_2(\epsilon)} = 1 \cdot 1 = 1$$

so that $f_1 f_2 \simeq g_1 g_2$. □

Since $\simeq$ is an equivalence relation, we can partition $C^0[0,*]$ into equivalence classes, and denote the corresponding collection of equivalence classes by $\mathcal{C}$. In particular, there is a unique equivalence class containing the constant function, 0. For convenience, we will denote this class (and any member function) by 0, as well. Notice that if $f \not\simeq 0$, then $f(\epsilon) > 0$ for $\epsilon \succ 0$.

Lemma 7.2 d) says that multiplication is a well-defined operation on $\mathcal{C}$. To perform addition, subtraction, or division on equivalence classes, we must restrict attention to functions which are "nice" enough. A standard restriction is to look at only those functions

which "look like" exponentials, i.e., those classes which contain an exponential of the form $c\epsilon^r \in C^0[0,*]$ for $c, r \geq 0$. Thus, we define the set of *exponentially convergent* (Young, 1993) functions, $\mathcal{C}^+ \subset \mathcal{C}$ as those equivalence classes containing $c\epsilon^r$ for some $r, c \geq 0$. Intuitively, we want to focus on the collection of functions $\left\{ f \in C^+[0,*] \mid \exists\, r, c \geq 0,\ f \simeq c\epsilon^r \right\}$. For simplicity, we will blur the distinction between an equivalence class in $\mathcal{C}^+$ and its member functions. Likewise, we will abuse notation slightly and write $f \in \mathcal{C}^+$ instead of $f \in C^+[0,*]$ and $f \simeq c\epsilon^r$, for some $r, c \geq 0$. For example, we may observe that, as constant functions, $\mathbb{R}^+ \subset \mathcal{C}^+$.

Theorem 7.3. *There exist functions $R : \mathcal{C}^+ \to [0,\infty]$ and $C : \mathcal{C}^+ \to [0,\infty)$, such that, for all $f, g \in \mathcal{C}^+$:*

a) $f \simeq c\epsilon^r \not\simeq 0$ iff $C(f) = c > 0$ and $R(f) = r < \infty$, and

b) $f \simeq 0$ iff $C(f) = 0$ iff $R(f) = \infty$

Moreover:

a) for $f, g \in \mathcal{C}^+$, $f \simeq g$ iff $C(f) = C(g)$ and $R(f) = R(g)$;

b) if f is continuous for $\epsilon \succeq 0$ and $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} = 1$, with $c > 0$, then $f \in \mathcal{C}^+ \setminus 0$ with $R(f) = r$ and $C(f) = c$.

Proof. First, observe that the mapping $\alpha : (0,\infty) \times [0,\infty) \to \mathcal{C}^+$ such that $\alpha(c,r) = c\epsilon^r$ gives a 1-1 correspondence between $(0,\infty) \times [0,\infty)$ and $\mathcal{C}^+ \setminus 0$. Assume that $\alpha(c_1, r_1) = \alpha(c_2, r_2)$, that is $c_1\epsilon^{r_1} \simeq c_2\epsilon^{r_2}$. Since $c_i > 0$, $i = 1, 2$, we are in case ii) of Definition 7.1. Therefore, $1 = \lim_{\epsilon \to 0^+} \frac{c_1\epsilon^{r_1}}{c_2\epsilon^{r_2}} = \frac{c_1}{c_2} \lim_{\epsilon \to 0^+} \epsilon^{r_1 - r_2}$. If $r_1 > r_2$ this limit is 0. If $r_1 < r_2$, the limit is ∞. Thus, we must have $r_1 = r_2$. Moreover, $1 = \frac{c_1}{c_2} \lim_{\epsilon \to 0^+} \epsilon^0 = \frac{c_1}{c_2}$ and $c_1 = c_2$. Thus, α is 1-1.

By definition, if $f \in \mathcal{C}^+ \setminus 0$, then $f \simeq c\epsilon^r$ for some $c, r \geq 0$. Since $f \not\simeq 0$, we must have $c > 0$, so that $f \simeq \alpha(c,r)$. Thus, α maps onto $\mathcal{C}^+ \setminus 0$. In particular, there are unique functions, R and C, such that $(R, C) : \mathcal{C}^+ \setminus 0 \to (0,\infty) \times [0,\infty)$ is the inverse of α. Notice that, if $f \simeq c\epsilon^r \not\simeq 0$, then $0 < c = C(\alpha(c,r)) = C(c\epsilon^r) = C(f)$ and $\infty > r = R(\alpha(c,r)) = R(c\epsilon^r) = R(f)$. Conversely, if $0 < c = C(f)$ and $\infty > r = R(f)$, then $f \equiv \alpha(C(f), R(f)) = c\epsilon^r \not\simeq 0$.

We may extend both functions to all of $\mathcal{C}^+$ by setting $C(0) = 0$ and $R(0) = \infty$. Thus, if $f \simeq 0$ then $C(f) = 0$ and $R(f) = \infty$. Conversely, if $C(f) = 0$ or $R(f) = \infty$, we must have $f \simeq 0$, since $C(f) > 0$ and $R(f) < \infty$ on $\mathcal{C}^+ \setminus 0$.

Next, observe that, since C and R are defined on equivalence classes, if $f \simeq g$, then necessarily $C(f) = C(g)$ and $R(f) = R(g)$. Conversely, if $C(f) = C(g)$ and $R(f) = R(g)$, we may show that $f \simeq g$. If $C(f) = C(g) = 0$, then $f \simeq 0 \simeq g$. Otherwise, $C(f) = C(g) > 0$, so that $R(f) = R(g) < \infty$, and $f \simeq C(f)\epsilon^{R(f)} = C(g)\epsilon^{R(g)} \simeq g$.

Finally, assume that f is continuous for $\epsilon \succeq 0$ and $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} = 1$. Then we must have $\frac{f(\epsilon)}{c\epsilon^r}$, and hence $f(\epsilon)$, be positive for $\epsilon \succ 0$. In particular, $f \in \mathcal{C} \setminus 0$. However, by assumption $f \simeq c\epsilon^r \not\simeq 0$, so by part a), $C(f) = c$ and $R(f) = r$. □

By our comments preceding Theorem 7.3, we can and will also view $C(f)$ and $R(f)$ as functions defined for all $f \in C^+[0, *]$ with $f \simeq c\epsilon^r$, for some $r, c \geq 0$, which are constant on equivalence classes.

We call the functions, $R(f)$ and $C(f)$, of Theorem 7.3, the *resistance* and *communication cost*, respectively, of f. The following Lemma shows how the resistance and communication cost functions behave with respect to addition, subtraction, multiplication, division, and taking limits.

Theorem 7.4. *The following hold for any $f, g \in \mathcal{C}^+$.*

a) $\lim_{\epsilon \to 0^+} f(\epsilon) = [R(f) = 0]C(f)$.

b) $f + g \in \mathcal{C}^+$, *with* $R(f + g) = \min\{R(f), R(g)\}$ *and* $C(f + g) = [R(f + g) = R(f)]C(f) + [R(f + g) = R(g)]C(g)$.

c) *If* $R(f) < R(g)$, *or* $R(f) = R(g)$ *and* $C(f) > C(g)$, *then* $f - g \in \mathcal{C}^+$, $C(f - g) = C(f) - [R(f) = R(g)]C(g)$, *and* $R(f - g) = R(f)$.

d) $fg \in \mathcal{C}^+$, *with* $C(fg) = C(f)C(g)$ *and* $R(fg) = R(f) + R(g)$.

e) *If* $g \not\simeq 0$, $R(f) \geq R(g)$, *and we define* $\left(\frac{f}{g}\right)(0) = [R(f) = R(g)]\frac{C(f)}{C(g)}$, *then* $\frac{f}{g} \in \mathcal{C}^+$, $C\left(\frac{f}{g}\right) = \frac{C(f)}{C(g)}$, *and* $R\left(\frac{f}{g}\right) = R(f) - R(g)$.

Proof. We first prove part a). If $f \not\simeq 0$, by Theorem 7.3 a), since $f \in \mathcal{C}^+$, $f \simeq C(f)\epsilon^{R(f)}$. Therefore,

$$C(f)[R(f) = 0] = \lim_{\epsilon \to 0^+} C(f)\epsilon^{R(f)} = \lim_{\epsilon \to 0^+} \frac{C(f)\epsilon^{R(f)}}{f(\epsilon)} \lim_{\epsilon \to 0^+} f(\epsilon) = \lim_{\epsilon \to 0^+} f(\epsilon)$$

Otherwise, $f \simeq 0$, $R(f) = \infty$, $C(f) = 0$, and $[R(f) = 0]C(f) = 0 = \lim_{\epsilon \to 0^+} f(\epsilon)$.

We will approach the proof of part b) in by cases. First, assume that either $f \simeq 0$ or $g \simeq 0$. Since the statement of part b) is symmetric in f and g, we may assume, without loss of generality, that $g \simeq 0$. Thus, $g(\epsilon) = 0$ and $f(\epsilon) + g(\epsilon) = f(\epsilon)$, for $\epsilon \succeq 0$, so that, by Lemma 7.2 a), $f + g \simeq f \in \mathcal{C}^+$. Moreover, by Theorem 7.3 a), $R(f+g) = R(f) = \min\{R(f), \infty\} = \min\{R(f), R(g)\}$ and $C(f+g) = C(f) = [R(f+g) = R(f)]C(f) + [R(f+g) = R(g)]C(g)$, as desired.

To complete the proof of part b), we may then assume that $f \not\simeq 0$ and $g \not\simeq 0$. In particular, we know that $f(\epsilon), g(\epsilon) > 0$ for $\epsilon \succ 0$, $C(f), C(g) > 0$, and $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{C(f)\epsilon^{R(f)}} = 1 = \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}$. By symmetry, we may assume that $R(g) \leq R(f)$.

Letting $c = [R(f) = R(g)]C(f) + C(g)$, we have

$$\begin{aligned}
\lim_{\epsilon \to 0^+} \frac{f(\epsilon) + g(\epsilon)}{c\epsilon^{R(g)}} &= \frac{1}{c}\left(C(f) \lim_{\epsilon \to 0^+} \frac{f(\epsilon)\epsilon^{R(f)-R(g)}}{C(f)\epsilon^{R(f)}} + C(g) \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}\right) \\
&= \frac{1}{c}\left(C(f) \cdot 1 \cdot \lim_{\epsilon \to 0^+} \epsilon^{R(f)-R(g)} + C(g) \cdot 1\right) \\
&= \frac{1}{c}\left(C(f)[R(f) = R(g)] + C(g)\right) = 1
\end{aligned}$$

Thus, we have shown that $f + g \simeq ([R(f) = R(g)]C(f) + C(g))\epsilon^{R(g)}$. In particular, $f + g \in \mathcal{C}^+$. Since $f(\epsilon) + g(\epsilon) > 0$ for $\epsilon \succ 0$, $f + g \not\simeq 0$, so by Theorem 7.3 a), we may conclude that $R(f+g) = R(g) = \min\{R(g), R(f)\}$ and $C(f+g) = [R(f) = R(g)]C(f) + C(g) = [R(f+g) = R(f)]C(f) + [R(f+g) = R(g)]C(g)$, as desired.

Now consider part c). First, consider the case when $g \simeq 0$, so that $R(g) = \infty$, $C(g) = 0$, $g(\epsilon) = 0$, for $\epsilon \succeq 0$. Therefore, $f(\epsilon) - g(\epsilon) = f(\epsilon)$, for $\epsilon \succeq 0$ and, by Lemma 7.2 a), $f - g \simeq f$. Thus, $R(f-g) = R(f)$ and, since $C(g) = 0$, $C(f-g) = C(f) = C(f) - [R(f) = R(g)]C(g)$, as desired.

Now assume that $g \not\simeq 0$, so that $C(f), C(g) > 0$. We know that $f(\epsilon) - g(\epsilon)$ is continuous for $\epsilon \succeq 0$, since both f and g are. Therefore, by Theorem 7.3 b), it only remains to calculate $R(f-g)$ and $C(f-g)$. As in the proof of part b), let $c = C(f) - [R(f) = R(g)]C(g)$. Either $R(f) < R(g)$, so $C(f) > 0$, or $R(f) = R(g)$, so that $C(f) > C(g)$. In both cases,

$c > 0$ and we may compute

$$\begin{aligned}\lim_{\epsilon\to 0^+}\frac{f(\epsilon)-g(\epsilon)}{c\epsilon^{R(f)}} &= \frac{1}{c}\left(C(f)\lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{C(f)\epsilon^{R(f)}}+C(g)\lim_{\epsilon\to 0^+}\frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}\epsilon^{R(g)-R(f)}\right)\\ &= \frac{1}{c}\left(C(f)\cdot 1+C(g)\cdot 1\cdot\lim_{\epsilon\to 0^+}\epsilon^{R(g)-R(f)}\right)\\ &= \frac{1}{c}\left(C(f)-[R(f)=R(g)]C(g)\right)=1\end{aligned}$$

Therefore, by Theorem 7.3 b), $f-g\in\mathcal{C}^+$, $C(f-g)=c=C(f)-[R(f)=R(g)]C(g)$, and $R(f-g)=R(f)$, as desired, completing the proof of part c).

To prove part d), by Lemma 7.2 b), we know that $fg\in\mathcal{C}$, so it remains to show that $fg\simeq c\epsilon^r$ for appropriately chosen r and c. First, consider the case when one of the factors, say, $f\simeq 0$. Then, $C(f)=0$, $R(f)=\infty$, and $f(\epsilon)\equiv 0$ for $\epsilon\succeq 0$. Thus, $f(\epsilon)g(\epsilon)\equiv 0$ for $\epsilon\succeq 0$ and $fg\simeq 0$. In particular, $fg\in\mathcal{C}^+$, with $C(fg)=0=0\cdot C(g)=C(f)C(g)$ and $R(fg)=R(0)=\infty=\infty+R(g)=R(f)+R(g)$ as desired.

Now assume that neither factor is 0, so that $C(f),C(g)>0$, $R(f),R(g)<\infty$, $f(\epsilon)g(\epsilon)>0$ for $\epsilon\succ 0$, and $\lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{C(f)\epsilon^{R(f)}}=1=\lim_{\epsilon\to 0^+}\frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}$. In this case,

$$\lim_{\epsilon\to 0^+}\frac{f(\epsilon)g(\epsilon)}{(C(f)C(g))\epsilon^{R(f)+R(g)}} = \lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{C(f)\epsilon^{R(f)}}\lim_{\epsilon\to 0^+}\frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}=1\cdot 1=1$$

Thus, $fg\simeq(C(f)C(g))\epsilon^{R(f)+R(g)}$, and $fg\in\mathcal{C}^+$. Since $fg\not\simeq 0$, the equations $C(fg)=C(f)C(g)$ and $R(fg)=R(f)+R(g)$ then follow directly from Theorem 7.3 a), as desired, completing the proof of part d).

Now to prove part e), we assume that $g\not\simeq 0$ and $R(f)\geq R(g)$. Thus, $C(g)>0$, $R(g)<\infty$, $g(\epsilon)>0$ for $\epsilon\succ 0$, and $\lim_{\epsilon\to 0^+}\frac{g(\epsilon)}{C(g)\epsilon^{R(g)}}=1$. First, consider the case when $f\simeq 0$, so that $C(f)=0$, $R(f)=\infty$, and $f(\epsilon)\equiv 0$ for $\epsilon\succeq 0$. Then $\frac{f(\epsilon)}{g(\epsilon)}\equiv 0$ for $\epsilon\succ 0$, $\lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{g(\epsilon)}=0$, and $\frac{f}{g}\in\mathcal{C}$ with $\left(\frac{f}{g}\right)(0)=0=[R(f)=R(g)]\frac{C(f)}{C(g)}$. Moreover, $\frac{f}{g}\simeq 0$, so that $\frac{f}{g}\in\mathcal{C}^+$, with $C\left(\frac{f}{g}\right)=0=\frac{0}{C(g)}=\frac{C(f)}{C(g)}$ and $R\left(\frac{f}{g}\right)=\infty=\infty-R(g)=R(f)-R(g)$, as desired.

Otherwise, $f\not\simeq 0$, so that $C(f)>0$, $R(f)<\infty$, $f(\epsilon)>0$ for $\epsilon\succ 0$, and $\lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{C(f)\epsilon^{R(f)}}=1$. Then

$$\begin{aligned}\lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{g(\epsilon))} &= \lim_{\epsilon\to 0^+}\frac{f(\epsilon)}{C(f)\epsilon^{R(f)}}\lim_{\epsilon\to 0^+}\frac{C(f)\epsilon^{R(f)}}{C(g)\epsilon^{R(g)}}\lim_{\epsilon\to 0^+}\frac{C(g)\epsilon^{R(g)}}{g(\epsilon)}\\ &= \lim_{\epsilon\to 0^+}\frac{C(f)\epsilon^{R(f)}}{C(g)\epsilon^{R(g)}}=\frac{C(f)}{C(g)}\lim_{\epsilon\to 0^+}\epsilon^{R(f)-R(g)}=[R(f)=R(g)]\frac{C(f)}{C(g)}\end{aligned}$$

Thus, setting $\left(\frac{f}{g}\right)(0) = [R(f) = R(g)]\frac{C(f)}{C(g)}$, $\frac{f}{g} \in C[0,*]$, and $\left(\frac{f}{g}\right)(\epsilon) > 0$ for $\epsilon \succ 0$, so that $\frac{f}{g} \in \mathcal{C} \setminus 0$. In addition,

$$\lim_{\epsilon \to 0^+} \frac{f(\epsilon)/g(\epsilon))}{C(f)\epsilon^{R(f)-R(g)}/C(g)} = \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{C(f)\epsilon^{R(f)}} \lim_{\epsilon \to 0^+} \frac{C(g)\epsilon^{R(g)}}{g(\epsilon)} = 1 \cdot 1 = 1$$

so that $\frac{f}{g} \simeq \frac{C(f)}{C(g)}\epsilon^{R(f)-R(g)}$. In particular, $\frac{f}{g} \in \mathcal{C}^+$ with $R\left(\frac{f}{g}\right) = R(f) - R(g)$, and $C\left(\frac{f}{g}\right) = \frac{C(f)}{C(g)}$, as desired. □

Parts b) and d) of Theorem 7.4 generalize to finite sums and products, as follows.

Corollary 7.5. *If $f_i \in \mathcal{C}^+$, $i = 1, \ldots, k$, then*

a) $f = \sum_{i=1}^k f_i \in \mathcal{C}^+$, *with* $R(f) = \min_{i \in S_k}\{R(f_i)\}$ *and* $C(f) = \sum_{i \in S_k}[R(f) = R(f_i)]C(f_i)$.

b) $f = \Pi_{i=1}^k f_i \in \mathcal{C}^+$, *with* $R(f) = \sum_{i \in S_k}\{R(f_i)\}$ *and* $C(f) = \Pi_{i \in S_k} C(f_i)$.

Proof. Both parts may be proven by induction. First, consider part a). The case when $k = 1$ is trivially true, since $f = f_1$. For $k > 1$, let $\overline{f} = \sum_{i \in S_{k-1}} f_i$ and apply Theorem 7.4 b) to $\overline{f}$ and f_k, along with the induction hypothesis, to obtain

$$R(f) = R\left(\overline{f} + f_k\right) = \min\left\{R\left(\overline{f}\right), R(f_k)\right\} = \min\left\{\min_{i \in S_{k-1}} R(f_i), R(f_k)\right\} = \min_{i \in S_k}\{R(f_i)$$

Likewise,

$$\begin{aligned} C(f) &= C\left(\overline{f} + f_k\right) = [R(f) = R\left(\overline{f}\right)]C\left(\overline{f}\right) + [R(f) = R(f_k)]C(f_k) \\ &= \sum_{i \in S_{k-1}} [R(f) = R\left(\overline{f}\right)][R\left(\overline{f} = R(f_i)\right)]C(f_i) + [R(f) = R(f_k)]C(f_k) \end{aligned}$$

Since we want this to equal $\sum_{i \in S_k}[R(f) = R(f_i)]C(f_i)$, it remains to show that $[R(f) = R\left(\overline{f}\right)][R\left(\overline{f} = R(f_i)\right)] = [R(f) = R(f_i)]$ for $i \in S_{k-1}$.

In general, since $[P][Q] = [P \text{ and } Q]$, this reduces to showing that, for any $i \in S_{k-1}$, $R(f) = R(f_i) \iff R(f) = R\left(\overline{f}\right)$ and $R\left(\overline{f}\right) = R(f_i)$. Equivalently, we must show that $R(f) = R(f_i) \iff R(f) = R\left(\overline{f}\right)$ and $R(f) = R(f_i)$. This is true iff $R(f) = R(f_i) \implies R(f) = R\left(\overline{f}\right)$, or equivalently, $R(f) = R(f_i) \implies R(f_i) = R\left(\overline{f}\right)$. Thus, by the formula for R given above, it suffices to observe that, for any $i \in S_{k-1}$, $R(f_i) = \min_{j \in S_k} R\left(f_j\right) \implies R(f_i) = \min_{j \in S_{k-1}} R\left(f_j\right)$.

The proof of part b) is a bit easier. The case of $k = 1$ is trivial. When $k > 1$, define $\overline{f} = \Pi_{i \in S_{k-1}} f_i$ and apply Theorem 7.4 b) to $\overline{f}$ and f_k, along with the induction hypothesis, to obtain

$$R(f) \;=\; R\left(\overline{f} f_k\right) = R\left(\overline{f}\right) + R(f_k) = \left(\sum_{i \in S_{k-1}} R(f_i)\right) + R(f_k) = \sum_{i \in S_k} R(f_i)$$

Likewise,

$$C(f) \;=\; C\left(\overline{f} f_k\right) = C\left(\overline{f}\right) C(f_k) = \left(\Pi_{i \in S_{k-1}} C(f_i)\right) C(f_k) = \Pi_{i \in S_k} C(f_i)$$

□

7.2 Perturbed Matrices

Before defining a PMM (perturbed *Markov* matrix), we first define simply a "perturbed matrix". Notice that Theorem 7.4 implies that $\mathcal{C}^+$ is closed under addition and multiplication. Thus, we may define a *perturbed matrix* as a matrix, $M_\epsilon \in \mathrm{Mat}\left(\mathcal{C}^+\right)$, that is, a matrix with entries in $\mathcal{C}^+$. As we mentioned in Section 7.1, by this we mean a matrix with entries in $C^+[0, *]$ which are exponentially convergent (i.e., whose equivalence class in $\mathcal{C}$ belongs to $\mathcal{C}^+$). Denoting the set of $n \times m$ perturbed matrices as $\mathrm{Pert}(n, m)$, and the set of all perturbed matrices by Pert, while subtraction and inversion are only defined in very limited circumstances, we will show that Pert is closed under addition and multiplication (assuming compatible dimensions).[2]

We begin by extending the definitions of R and C to Pert. For any perturbed matrix M_ϵ, we may define the associated *resistance* matrix, $R(M_\epsilon)$, where $R(M_\epsilon)_{i,j} = R\left((M_\epsilon)_{ij}\right)$. We likewise define its associated *cost* matrix, $C_\epsilon(M_\epsilon)$, where $C(M_\epsilon)_{i,j} = C\left((M_\epsilon)_{ij}\right)$.[3]

We will say that two perturbed matrices, M_ϵ and ${M'}_\epsilon$ are *asymptotically equal* and write $M_\epsilon \simeq {M'}_\epsilon$ iff $(M_\epsilon)_{ij} \simeq \left({M'}_\epsilon\right)_{ij}$ for all i, j. Notice that, by Theorem 7.3, $M_\epsilon \simeq {M'}_\epsilon$ iff $C(M_\epsilon) = C\left(M'_\epsilon\right)$ and $R(M_\epsilon) = R\left(M'_\epsilon\right)$.

Theorem 7.4 then generalizes as follows.

[2] Thus, we may also multiply by "scalars" in $\mathcal{C}^+$, since multiplication by $f \in \mathcal{C}^+$ is the same as multiplying by fI, the diagonal matrix with all diagonal entries equal to f

[3] In some contexts, this is also known as the *communication* matrix of M_ϵ.

Theorem 7.6. *Assume that* $M_\epsilon, M'_\epsilon \in \mathrm{Pert}(n,m)$*, while* $\widetilde{M}_\epsilon \in \mathrm{Pert}(m,p)$.

a) $M_\epsilon + M'_\epsilon \in \mathrm{Pert}(n,m)$

b) $C\left(M_\epsilon + M'_\epsilon\right)_{i,j} = \left[R\left(M_\epsilon + M'_\epsilon\right)_{i,j} = R\left(M_\epsilon\right)_{i,j}\right] C\left(M_\epsilon\right)_{i,j}$
$+ \left[R\left(M_\epsilon + M'_\epsilon\right)_{i,j} = R\left(M'_\epsilon\right)_{i,j}\right] C\left(M'_\epsilon\right)_{i,j}$

c) $R\left(M_\epsilon + M'_\epsilon\right)_{i,j} = \min\left\{R\left(M_\epsilon\right)_{i,j}, R\left(M'_\epsilon\right)_{i,j}\right\}$

d) $M_\epsilon \widetilde{M}_\epsilon \in \mathrm{Pert}(n,p)$

e) $C\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} = \sum_{k\in S_m} C\left(M'_\epsilon\right)_{i,k} C\left(\widetilde{M}_\epsilon\right)_{k,j} \left[R\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} = R\left(M_\epsilon\right)_{i,k} + R\left(\widetilde{M}_\epsilon\right)_{k,j}\right]$

f) $R\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} = \min_{k\in S_m}\left\{R\left(M_\epsilon\right)_{i,k} + R\left(\widetilde{M}_\epsilon\right)_{k,j}\right\}$

In particular, addition and multiplication of perturbed matrices is well-defined on equivalence classes under $\simeq$.

Proof. Parts a), b), c) and d) follow immediately from the definitions and parts b) and d) of Theorem 7.4. To prove part e), apply Corollary 7.5 and Theorem 7.4 d):

$$\begin{aligned} C\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} &= C\left(\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j}\right) = C\left(\sum_{k\in S_m} \left(M_\epsilon\right)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right) \\ &= \sum_{k\in S_m} C\left(\left(M'_\epsilon\right)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right) \left[R\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} = R\left(\left(M_\epsilon\right)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right)\right] \\ &= \sum_{k\in S_m} C\left(M'_\epsilon\right)_{i,k} C\left(\widetilde{M}_\epsilon\right)_{k,j} \left[R\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} = R\left(M_\epsilon\right)_{i,k} + R\left(\widetilde{M}_\epsilon\right)_{k,j}\right] \end{aligned}$$

Similarly, by Corollary 7.5 and Theorem 7.4 d)

$$\begin{aligned} R\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j} &= R\left(\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j}\right) = R\left(\sum_{k\in S_m} \left(M_\epsilon\right)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right) \\ &= \min_{k\in S_m} R\left(\left(M'_\epsilon\right)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right) \\ &= \min_{k\in S_m}\left\{R\left(M'_\epsilon\right)_{i,k} + R\left(\widetilde{M}_\epsilon\right)_{k,j}\right\} \end{aligned}$$

□

7.3 Perturbed Markov Matrices and Stable Distributions

In this section, we formally define what we mean by a "perturbed" Markov matrix, M_ϵ, and all the associated concepts from Part I. That is, we define

- the weighted and unweighted graphs associated with M_ϵ,
- the additional graphs associated with the "unperturbed" Markov matrix, M_0,
- the stable and stochastically stable distributions of M_ϵ, and
- the collections of rooted, directed spanning subtrees associated with M_ϵ with their corresponding weight functions.

The most delicate issue in this section is to prove that the stable distribution is sufficiently well-behaved (i.e., is a perturbed matrix) so that we may take its limit as $\epsilon \to 0$ to even define its stochastically stable distribution. This will involve a careful application of the Markov Chain Tree Theorem from chapter 6, where we will define both the "resistance" and "cost" of a subtree, and restrict attention to minimal weight directed, spanning subtrees.

We now define a *perturbed Markov matrix* (PMM) as a perturbed matrix M_ϵ such that, for $\epsilon \succeq 0$, M_ϵ is a Markov matrix and is unichain for $\epsilon \succ 0$.[4] Notice that to say that M_ϵ is Markov is equivalent to saying that $(M_\epsilon)_{j,j} = 1 - \sum_{i \neq j} (M_\epsilon)_{i,j} \in \mathcal{C}^+$. Since $\mathcal{C}^+$ is not closed under subtraction, in general, this is a somewhat subtle assumption. We will denote the set of $n \times n$ perturbed Markov matrices by $\mathrm{PMM}(n)$. We define its associated *perturbed graph*, as a weighted, directed graph, but where the weight on each edge is in $\mathcal{C}^+$. Formally, $G(M_\epsilon) = (V, E, d)$, so that $V = \{v_1, \ldots, v_n\}$, with $(v_i, v_j) \in E$ iff $(M_\epsilon)_{j,i} \not\approx 0$, and $d(v_i, v_j) = (M_\epsilon)_{j,i}$. Notice, in particular, that the graph does not contain edge (v_i, v_j) iff $R(M_\epsilon)_{j,i} = \infty$, corresponding to the intuition that current does not flow through a wire with "infinite" resistance. As before, we will denote the underlying unweighted graph as $G_-(M_\epsilon)$, and its transitive closure by $\mathcal{P}(M_\epsilon) \equiv (G_-(M_\epsilon))_T$. Remember that (v_i, v_j) is an edge in $\mathcal{P}(M_\epsilon)$ iff there is a walk from v_i to v_j in $G_-(M_\epsilon)$ iff there is a path from v_i to v_j in $G_-(M_\epsilon)$. Thus, we may call this the "path" graph of M_ϵ.

As before, we define strongly connected components, closed classes, invariant and transient sets of indices in terms of the corresponding collection of vertices in $G_-(M_\epsilon)$. We

[4] This generalizes the usual definition of a perturbed Markov matrix, which requires that M_ϵ be irreducible.

should point out that the unweighted graphs corresponding to M_ϵ for each $\epsilon \succ 0$ (which we would also denote by $G_-(M_\epsilon)$), are all the same (by definition of $C^+[0,*]$) and equal to $G_-(M_\epsilon)$.[5] That is, although we could interpret the notation $G_-(M_\epsilon)$ in two ways, either interpretation leads to exactly the same unweighted graph. Thus, for example, the closed classes of the perturbed matrix are just the same as those of the Markov matrix at any fixed, sufficiently small ϵ.

Moreover, we define the associated *unperturbed* graph, $G_0(M_\epsilon) = G(M_0)$, i.e., the Markov graph on the (unperturbed) Markov matrix, M_0. Notice that, since $c \cdot 0^r = c > 0$ iff $r = 0$ (and $c > 0$), $G(M_0)$ may also be thought of as the "zero-resistance" subgraph of $G(M_\epsilon)$, that is, the weighted directed subgraph of $G(M_\epsilon)$ containing all edges, (v_i, v_j), with $R\left(d\left(v_i, v_j\right)\right) = 0$ and edge weights given by $C\left(d\left(v_i, v_j\right)\right)$.

Let $\mathrm{stab}(M_\epsilon)$ denote the unique stable distribution of M_ϵ for $\epsilon \succ 0$ given by Corollary 5.15.[6] Using the notation of Theorem 6.17, $\mathrm{stab}(M_\epsilon) = v_{M_\epsilon}$ for $\epsilon \succ 0$. We will show that $\mathrm{stab}(M_\epsilon)$ may be defined at $\epsilon = 0$ so that its entries are all in $\mathcal{C}^+$. In particular, $\mathrm{ssd}(M_\epsilon) \equiv \lim_{\epsilon \to 0^+} \mathrm{stab}(M_\epsilon)$ exists. We call this limit the *stochastically stable distribution* of M_ϵ.

To prove this, we will need to extend the notation of Chapter 6. Since $G(M_\epsilon)$ has weights in $\mathcal{C}^+$, we have three notions of the total weight of a directed subtree. For any directed spanning tree rooted at i corresponding to $\sigma \in \overline{T}_i$, the total weight of σ in M_ϵ may be defined, just as in Chapter 6, as $W(M_\epsilon, \sigma) = \Pi_{j \neq i} (M_\epsilon)_{\sigma(j),j}$. By Corollary 7.5 b), $W(M_\epsilon, \sigma) \in \mathcal{C}^+$. Thus, we may also define the *resistance of σ in M_ϵ* as $R(M_\epsilon, \sigma) \equiv R(W(M_\epsilon, \sigma))$. Similarly, we define the *cost of σ in M_ϵ* as $C(M_\epsilon, \sigma) \equiv C(W(M_\epsilon, \sigma))$. By Corollary 7.5 b), the resistance and cost of the tree, σ, satisfy the following equations:

$$R(M_\epsilon, \sigma) = R(W(M_\epsilon, \sigma)) = R\left(\Pi_{j \neq i} (M_\epsilon)_{\sigma(j),j}\right) = \sum_{j \neq i} R\left((M_\epsilon)_{\sigma(j),j}\right) \quad (7.1)$$

$$C(M_\epsilon, \sigma) = C(W(M_\epsilon, \sigma)) = C\left(\Pi_{j \neq i} (M_\epsilon)_{\sigma(j),j}\right) = \Pi_{j \neq i} C\left((M_\epsilon)_{\sigma(j),j}\right) \quad (7.2)$$

In particular, since $\sigma \in \overline{T}_i$, $\sigma(j) \neq j$ for $j \neq i$, so that $R(M_\epsilon, \sigma)$ and $C(M_\epsilon, \sigma)$ do not depend on the diagonal entries entries of M_ϵ.

[5] In fact, this is the main reason why we define perturbed matrices in terms of $C^+[0,*]$.

[6] This is a slight abuse of notation, since, up to this point, stab would have referred to the singleton set containing the stable distribution.

Notice that, if $G_{-}(\sigma)$ is not a subgraph of $G_{-}(M_\epsilon)$, it does not contain some edge, $\left(v_{\sigma(j)}, v_j\right)$, so that $(M_\epsilon)_{\sigma(j),j} \simeq 0$. In terms of the resistance, the resistance of such an edge is ∞, so that the sum, $R(M_\epsilon, \sigma) = \infty$. Conversely, if the sum is infinite, the resistance of some edge is infinite, implying σ is not a subgraph of $G_{-}(M_\epsilon)$. Likewise, $C(M_\epsilon, \sigma) = 0$ iff σ is not a subgraph of $G_{-}(M_\epsilon)$. Thus, we may define the set of directed spanning subtrees of $G_{-}(M_\epsilon)$, rooted at i, as

$$T(M_\epsilon, i) \equiv \left\{\sigma \in \overline{T}_i \mid R(M_\epsilon, \sigma) < \infty\right\} = \left\{\sigma \in \overline{T}_i \mid C(M_\epsilon, \sigma) > 0\right\}$$

As before, let $T(M_\epsilon) \equiv \bigcup_{i \in S_n} T(M_\epsilon, i)$.

We now prove the statement above regarding $\operatorname{stab}(M_\epsilon)$. Moreover, we give formulas for its resistance and cost in terms of the resistance and cost of the directed spanning subtrees of $G_{-}(M_\epsilon)$.

Theorem 7.7. *If $M_\epsilon \in \mathrm{PMM}(n)$, if we define*

$$\begin{aligned} r_i &\equiv \min_{\sigma \in T(M_\epsilon, i)} R(M_\epsilon, \sigma) \text{ and } r \equiv \min_{\sigma \in T(M_\epsilon)} R(M_\epsilon, \sigma) \\ \overline{T}(M_\epsilon, i) &\equiv \left\{\sigma \in \overline{T}_i \mid R(M_\epsilon, \sigma) = r_i\right\} \text{ and } \overline{T}(M_\epsilon) \equiv \left\{\sigma \in \overline{T}_i \mid i \in S_n, R(M_\epsilon, \sigma) = r\right\} \end{aligned}$$

then

a) there exists a perturbed column vector, $\operatorname{stab}(M_\epsilon) \in \mathrm{Pert}(n, 1)$, which, for each $\epsilon \succ 0$, is the unique stable distribution of M_ϵ,

b) $r_i = \min_{\sigma \in \overline{T}_i} R(M_\epsilon, \sigma)$ and $r = \min_i r_i$,

c) $R(\operatorname{stab}(M_\epsilon))_i = r_i - r$, and

d) $C(\operatorname{stab}(M_\epsilon))_i = \frac{\sum_{\sigma \in \overline{T}(M_\epsilon, i)} C(M_\epsilon, \sigma)}{\sum_{\sigma \in \overline{T}(M_\epsilon)} C(M_\epsilon, \sigma)}$

Proof. Using the notation of Theorem 6.17, fix a perturbed Markov matrix, M_ϵ, and abbreviate w_{M_ϵ} as w_ϵ. We first show that $w_{M_\epsilon} \in \mathrm{Pert}(n, 1)$. By Equation 6.4, $(w_\epsilon)_i = \sum_{\sigma \in T(M_\epsilon, i)} W(M_\epsilon, \sigma)$. Therefore, by Corollary 7.5, $(w_\epsilon)_i \in \mathcal{C}^+$ and

$$R((w_\epsilon)_i) = R\left(\sum_{\sigma \in T(M_\epsilon, i)} W(M_\epsilon, \sigma)\right) = \min_{\sigma \in T(M_\epsilon, i)} R(M_\epsilon, \sigma) = r_i$$

Since $R(M_\epsilon,\sigma)=\infty$, for any $\sigma\in\overline{T}_i\backslash T(M_\epsilon,i)$, we also have $r_i=\min_{\sigma\in T(M_\epsilon,i)}R(M_\epsilon,\sigma)=\min_{\sigma\in\overline{T}_i}R(M_\epsilon,\sigma)$. Moreover, since $T(M_\epsilon)=\bigcup_i T(M_\epsilon,i)$,

$$r=\min_{\sigma\in T(M_\epsilon)}R(M_\epsilon,\sigma)=\min_i\min_{\sigma\in T(M_\epsilon,i)}R(M_\epsilon,\sigma)=\min_i r_i$$

Set $K_\epsilon\equiv Jw_\epsilon$. Then $K_\epsilon\in\mathcal{C}^+$ and

$$K_\epsilon=\sum_i(w_\epsilon)_i=\sum_i\sum_{\sigma\in T(M_\epsilon,i)}W(M_\epsilon,\sigma)=\sum_{\sigma\in T(M_\epsilon)}W(M_\epsilon,\sigma)$$

Corollary 7.5 again gives

$$R(K_\epsilon)=R\left(\sum_{\sigma\in T(M_\epsilon)}W(M_\epsilon,\sigma)\right)=\min_{\sigma\in T(M_\epsilon)}R(M_\epsilon,\sigma)=r$$

Since M_ϵ is unichain, $K_\epsilon>0$ for $\epsilon\succ 0$, so that $r<\infty$. Moreover, since $r\leq r_i$, Theorem 7.4 e) implies that $\frac{w_\epsilon}{K_\epsilon}\in\mathrm{Pert}(n,1)$.

Thus, we may define $\mathrm{stab}(M_\epsilon)=\frac{w_\epsilon}{K_\epsilon}$. For $\epsilon\succ 0$, by Theorem 6.17, we then have $v_{M_\epsilon}=\frac{w_\epsilon}{Jw_\epsilon}=\mathrm{stab}(M_\epsilon)$. Moreover, by Theorem 7.4 e), $R(\mathrm{stab}(M_\epsilon)_i)=R(w_\epsilon)-R(K_\epsilon)=r_i-r$.

By Corollary 7.5 a),

$$\begin{aligned}C((w_\epsilon)_i) &= C\left(\sum_{\sigma\in T(M_\epsilon,i)}W(M_\epsilon,\sigma)\right)=\sum_{\sigma\in T(M_\epsilon,i)}[r_i=R(M_\epsilon,\sigma)]\,C(M_\epsilon,\sigma)\\ &= \sum_{\sigma\in\overline{T}(M_\epsilon,i)}C(M_\epsilon,\sigma)\end{aligned}$$

Likewise,

$$\begin{aligned}C(K_\epsilon) &= C\left(\sum_{\sigma\in T(M_\epsilon)}W(M_\epsilon,\sigma)\right)=\sum_{\sigma\in T(M_\epsilon)}[r=R(M_\epsilon,\sigma)]\,C(M_\epsilon,\sigma)\\ &= \sum_{\sigma\in\overline{T}(M_\epsilon)}C(M_\epsilon,\sigma)\end{aligned}$$

Thus, by Theorem 7.4 e), $C(\mathrm{stab}(M_\epsilon)_i)=\frac{C(w_\epsilon)}{C(K_\epsilon)}=\frac{\sum_{\sigma\in\overline{T}(M_\epsilon,i)}C(M_\epsilon,\sigma)}{\sum_{\sigma\in\overline{T}(M_\epsilon)}C(M_\epsilon,\sigma)}$. □

Example 7.8. To illustrate the ideas of Theorem 7.7, consider

$$M_\epsilon = \begin{pmatrix} 1-\epsilon-\epsilon^2 & \epsilon & \epsilon \\ \epsilon & 1-\epsilon-\epsilon^2 & \epsilon^3 \\ \epsilon^2 & \epsilon^2 & 1-\epsilon-\epsilon^3 \end{pmatrix}$$

then

$$\Lambda_\epsilon = M_\epsilon - I = \begin{pmatrix} -\epsilon-\epsilon^2 & \epsilon & \epsilon \\ \epsilon & -\epsilon-\epsilon^2 & \epsilon^3 \\ \epsilon^2 & \epsilon^2 & -\epsilon-\epsilon^3 \end{pmatrix}$$

Let's compute $(w_\epsilon)_1$. First way:

$$(w_\epsilon)_1 = \det \begin{vmatrix} -\epsilon-\epsilon^2 & \epsilon^3 \\ \epsilon^2 & -\epsilon-\epsilon^3 \end{vmatrix} = \epsilon^2+\epsilon^3+\epsilon^4$$

Second way: Find the set of mappings $\sigma : \{1,2,3\} - \{1\} \to \{1,2,3\}$ s. t. σ specifies the parent relation of a directed , spanning tree on $\{1,2,3\}$. In particular, $\sigma(i) \neq i$ (i.e., no loops), for all $i \in \{2,3\}$. There are three such mappings: $2 \mapsto 1, 3 \mapsto 1$; $2 \mapsto 3, 3 \mapsto 1$; $2 \mapsto 1, 3 \mapsto 2$.

$$(w_\epsilon)_1 = \sum_\sigma \prod_{p \neq 1} M_{\sigma(p),p} = (M_\epsilon)_{12}(M_\epsilon)_{13} + (M_\epsilon)_{32}(M_\epsilon)_{13} + (M_\epsilon)_{12}(M_\epsilon)_{23} = \epsilon^2+\epsilon^3+\epsilon^4$$

Likewise, $(w_\epsilon)_2 = \epsilon^2 + 2\epsilon^4$ and $(w_\epsilon)_3 = \epsilon^2 + 2\epsilon^3$. Thus, $m = 2$ and $u_0 = (1,1,1)^t$. The entries of u_0 correspond to the ϵ^2 terms in w_ϵ, which in turn correspond to the directed spanning trees given by: $2 \mapsto 1, 3 \mapsto 1$; $3 \mapsto 1, 1 \mapsto 2$; $2 \mapsto 1, 1 \mapsto 3$.

□

By Theorem 7.7, for each index, $i \in S_n$, $R(\mathrm{stab}(M_\epsilon))_i$ is well-defined. We will call this the *virtual energy*[7] at i in M_ϵ. By Theorem 7.4 a), we have that $(v_0(M_\epsilon))_i > 0$ iff $R(\mathrm{stab}(M_\epsilon))_i = 0$. Since, as we have seen in section 5.3, indices correspond to states in a Markov process, the collection of such indices are called the *stochastically stable* or *ground* states of M_ϵ, since these are precisely the indices for which the stochastically stable distribution, $v_0(M_\epsilon)$, has a non-zero component, and the stable distribution $\mathrm{stab}(M_\epsilon)$ has zero virtual energy (i.e., resistance).

[7] Desai, et. al. (Desai et al., 1994) call this the "stationary order" at i.

7.4 Equivalence of PMMs and Scaling

Since we are primarily focused, at present, on computing the stochastically stable distribution (SSD) of a PMM, we introduce an equivalence relation on PMMs for which the SSD is invariant. Likewise, we generalize the notion of a D-equivalence from chapter 5 to PMMs, where D may itself be a perturbed matrix. As in the non-perturbed case, such an equivalence will allow us to determine the SSD of a PMM from the SSD of an D_ϵ-equivalent one. Specifically, we will

- define an equivalence relation on PMMs and show that equivalent PMMs have equal SSDs,
- observe that asymptotically equal PMMs are equivalent,
- generalize the scaling construction from chapter 5 in two rather specific settings ("uniform" and "non-uniform" scaling), which will be crucial to our main algorithm in section 7.7,
- prove that these two constructions lead to equivalent or D_ϵ-equivalent results, respectively, and
- show these two constructions always guarantee "progress" in our algorithm to a solution.

Because scaling is defined in terms of subtraction and division, these results are rather delicate, since these operations are not generally defined in Pert.

To begin, we say that two perturbed Markov matrices are equivalent if they have asymptotically equal stable distributions. To state this formally,

Definition 7.9. *Two perturbed Markov matrices, M_ϵ and M'_ϵ are* equivalent, *denoted by $M_\epsilon \sim M'_\epsilon$, iff $\mathrm{stab}\,(M_\epsilon) \simeq \mathrm{stab}\left(M'_\epsilon\right)$.*

For example, we can show that asymptotically equal PMMs are equivalent.

Theorem 7.10. *If $M_\epsilon \simeq M'_\epsilon$, then $M_\epsilon \sim M'_\epsilon$.*

Proof. This follows from Theorem 7.7, Equations 7.1 and 7.2, and Theorem 7.3. Specifically, for $M_\epsilon, M'_\epsilon \in \mathrm{Pert}(n)$, to show that $M_\epsilon \sim M'_\epsilon$, by Theorem 7.3, we must show that

$R\left(\text{stab}\left(M_\epsilon\right)\right)_i = R\left(\text{stab}\left(M'_\epsilon\right)\right)_i$ and $C\left(\text{stab}\left(M_\epsilon\right)\right)_i = C\left(\text{stab}\left(M'_\epsilon\right)\right)_i$ for $i \in S_n$. By Theorem 7.7, since r_i and r can be defined with respect to $\overline{T}_i$, it suffices to show that for all $\sigma \in \overline{T}_i$, $R\left(M_\epsilon, \sigma\right) = R\left(M'_\epsilon, \sigma\right)$ and $C\left(M_\epsilon, \sigma\right) = C\left(M'_\epsilon, \sigma\right)$. By Equations 7.1 and 7.2 the resistance and cost of a tree depend only on the resistance and cost matrices. But, again by Theorem 7.3, since $M_\epsilon \simeq M'_\epsilon$, $R\left(M_\epsilon\right) = R\left(M'_\epsilon\right)$ and $C\left(M_\epsilon\right) = C\left(M'_\epsilon\right)$, so that $R\left(M_\epsilon, \sigma\right) = R\left(M'_\epsilon, \sigma\right)$ and $C\left(M_\epsilon, \sigma\right) = C\left(M'_\epsilon, \sigma\right)$ for all σ. □

When a uniform scaling of a PMM, M_ϵ, by $f \in \mathcal{C}^+$ yields another PMM, M'_ϵ, they are equivalent. That is, we have an analog of Lemma 5.3 for PMMs.

Theorem 7.11. *Given $f \in \mathcal{C}^+ \setminus 0$ and $M_\epsilon \in \text{PMM}(n)$, such that*

a) $R(f) \leq R\left(M_\epsilon\right)_{i,j}$ *for all* $i \neq j$,

b) $\left(M_\epsilon\right)_{j,j} + f(\epsilon) - 1 \in \mathcal{C}^+$, *and*

c) $R(f) \leq R\left(\left(M_\epsilon\right)_{j,j} + f(\epsilon) - 1\right)$ *for all* j,

if we define $M'_\epsilon = \frac{1}{f}\left(M_\epsilon - I\right) + I$, *then* $M'_\epsilon \in \text{PMM}(n)$ *and* $M_\epsilon \sim M'_\epsilon$.

Proof. The proof is similar to that of Lemma 5.3. The real work is in proving that $M'_\epsilon \in \text{PMM}(n)$. Since $f \not\simeq 0$ and $R(f) \leq R\left(M_\epsilon\right)_{i,j}$, we have, by Theorem 7.4 e), that $\left(M'_\epsilon\right)_{i,j} \in \mathcal{C}^+$ for all $i \neq j$. Moreover, $f(\epsilon) > 0$, $J\Lambda'_\epsilon = \frac{1}{f(\epsilon)} J\Lambda_\epsilon = 0$ for $\epsilon \succ 0$, and by continuity, also at $\epsilon = 0$. Thus, $\left(M'_\epsilon\right)_{j,j} = 1 - \sum_{i \neq j}\left(M'_\epsilon\right)_{i,j}$, for all $j \in S_n$ and $\epsilon \succeq 0$. Moreover, by Theorem 7.4 e), $\left(M'_\epsilon\right)_{j,j} = \frac{\left(M_\epsilon\right)_{j,j} + f - 1}{f} \in \mathcal{C}^+$. Thus, $\left(M'_\epsilon\right)_{i,j} \in \mathcal{C}^+$

It is now easy to show that $M_\epsilon \sim M'_\epsilon$. Since $f \not\simeq 0$, $f(\epsilon) > 0$, $\Lambda'_\epsilon = \frac{1}{f(\epsilon)}\Lambda_\epsilon$, and $\ker \Lambda'_\epsilon = \ker \Lambda_\epsilon$, so that $\text{stab}\left(M_\epsilon\right) = \text{stab}\left(M'_\epsilon\right)$ for $\epsilon \succ 0$. In particular, by Lemma 7.2 a), $\text{stab}\left(M_\epsilon\right) \simeq \text{stab}\left(M'_\epsilon\right)$, so that $M_\epsilon \sim M'_\epsilon$. □

In particular, when computing an SSD, we can always assume that $R\left(M_\epsilon\right)_{j,j} = 0$, for all j.

Corollary 7.12. *Given $M_\epsilon \in \text{PMM}(n)$, if we define $M'_\epsilon = \frac{1}{2}\left(M_\epsilon - I\right) + I$, then $M'_\epsilon \in \text{PMM}(n)$, $M_\epsilon \sim M'_\epsilon$, and $R\left(M'_\epsilon\right)_{j,j} = 0$, for all $j \in S_n$.*

Proof. While we could prove this directly fairly easily, we will instead appeal to Theorem 7.11 with $f(\epsilon) = 2 \not\simeq 0$. As are all positive, constant functions, $f \in \mathcal{C}+$, and by Theorem 7.4 a), $C(f) = f(0) = 2$ and $R(f) = 0$. Thus, it only remains to observe

that, by Theorem 7.4 b), $(M_\epsilon)_{j,j} + f(\epsilon) - 1 = (M_\epsilon)_{j,j} + 1 \in \mathcal{C}^+$. Since $R(f) = 0$, the remaining two conditions of the theorem follow immediately. Thus, by Theorem 7.11, $M'_\epsilon = \frac{1}{f}(M_\epsilon - I) + I = \frac{1}{2}(M_\epsilon - I) + I = \frac{1}{2}(M_\epsilon + I) \in \mathrm{PMM}(n)$, and $M_\epsilon \sim M'_\epsilon$. Moreover, by parts b) and e) of Theorem 7.4, $R\left(M'_\epsilon\right)_{j,j} = 0$, for all $j \in S_n$. □

Notice that equivalent PMMs must have equal virtual energies, since the energy at i is just the resistance of the i^{th} component of the stable distribution, and resistance is invariant under asymptotic equivalence. Likewise, their SSDs are equal, since, by Theorem 7.4 a), the components of their SSDs may be expressed a in terms of the resistance and cost of their respective stable distributions. Specifically, $\lim_{\epsilon\to 0^+} \mathrm{stab}(M_\epsilon)_i = [R(\mathrm{stab}(M_\epsilon)_i) = 0]C(\mathrm{stab}(M_\epsilon)_i)$, which is again is constant on asymptotic equivalence classes. In addition, they have the same path graphs.

Lemma 7.13. *Given $M_\epsilon, M'_\epsilon \mathrm{PMM}(n)$, if $M_\epsilon \sim M'_\epsilon$, then $\mathcal{P}(M_\epsilon) = \mathcal{P}\left(M'_\epsilon\right)$.*

Proof. By Theorems 6.8 and 6.17, since $\mathrm{stab}(M_\epsilon) \simeq \mathrm{stab}\left(M'_\epsilon\right)$,

$$s = \{i \in S_n \mid (\mathrm{stab}(M_\epsilon))_i \not\simeq 0\} = \left\{i \in S_n \mid \left(\mathrm{stab}\left(M'_\epsilon\right)\right)_i \not\simeq 0\right\} = s'$$

Since $\mathcal{P}(M_\epsilon)$ and $\mathcal{P}\left(M'_\epsilon\right)$ are both complete, directed graphs on their corresponding (unique) closed classes, V_s and $V_{s'}$, since $s = s'$, they must be equal. □

We also should generalize the notion of D-equivalence from section 5.1 to apply to PMMs.

Definition 7.14. *For $D_\epsilon \in \mathrm{Pert}$, we will say that two perturbed Markov matrices, M_ϵ and M'_ϵ, are D_ϵ-*equivalent*, and write $M'_\epsilon \sim_{D_\epsilon} M_\epsilon$, iff $M'_\epsilon \approx_{D_\epsilon} M_\epsilon$ for $\epsilon \succ 0$ and $R\left(\left\|D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)\right\|_1\right) = 0$.*

We then have the following analog to Lemma 5.2.

Theorem 7.15. *If $M_\epsilon \sim_{D_\epsilon} M'_\epsilon$, then $D_0 \mathrm{ssd}\left(M'_\epsilon\right) \propto \mathrm{ssd}(M_\epsilon)$.*

Proof. Since M_ϵ and M'_ϵ are both unichain, for $\epsilon \succ 0$, we must have $\mathrm{stab}(M_\epsilon) \propto D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)$, so that $f(\epsilon)\mathrm{stab}(M_\epsilon) = D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)$ for some function, f. In fact, since $f(\epsilon) = f(\epsilon)J\mathrm{stab}(M_\epsilon) = JD_\epsilon \mathrm{stab}\left(M'_\epsilon\right) = \left\|D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)\right\|_1$, which by Theorem 7.6 d) is in $Pert(1,1)$, i.e., $f \in \mathcal{C}^+$. By assumption, $R(f(\epsilon)) = 0$, so by Theorem 7.4 a), $C(f)\,\mathrm{ssd}(M_\epsilon) = D_0\,\mathrm{ssd}\left(M'_\epsilon\right)$. In particular, $D_0\,\mathrm{ssd}\left(M'_\epsilon\right) \propto \mathrm{ssd}(M_\epsilon)$. □

Our algorithm in section 7.7 requires that, if it is not unichain, M_0 should possess at least one non-trivial (i.e., containing more than one element) communicating class, which is not always the case for an arbitrary perturbed Markov matrix, M_ϵ. However, in this case we may transform M_ϵ to a closely related perturbed Markov matrix.

Lemma 7.16. *Given $M_\epsilon \, \mathrm{PMM}(n)$, such that all communicating classes of M_0 are singletons, define $s = \min_{j \notin \mathcal{T}} \min_{i \neq j} R\left(M_\epsilon\right)_{i,j}$, where $\mathcal{T}$ is the set of transient states of M_0. Likewise, let $c = 2 \max_{j \notin \mathcal{T}} \sum_{s=R(M_\epsilon)_{i,j}} C\left(M_\epsilon\right)_{i,j}$. If M_0 possesses more than one closed class or M_ϵ is irreducible with $n > 1$, then $0 < s < \infty$ and $c > 0$. In addition, if $f(\epsilon) = c\epsilon^s$ and*

$$\left(i_\epsilon\right)_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ f(\epsilon) & \text{if } i = j \text{ and } j \in \mathcal{T} \\ 1 & \text{otherwise} \end{cases}$$

then $i_\epsilon \in \mathrm{Pert}(n)$ and, using the notation of Chapter 5, $\overline{M}_\epsilon \equiv \left(M_\epsilon\right)_{i_\epsilon} \in \mathrm{PMM}(n)$ and $\overline{M}_\epsilon \approx_{i_\epsilon} M_\epsilon$, for $\epsilon \succ 0$. Moreover, f satisfies the assumptions of Lemma 7.11 with respect to $\overline{M}_\epsilon$, so that $M'_\epsilon = \frac{1}{f}\left(\overline{M}_\epsilon - I\right) + I$ is i_ϵ-equivalent to M_ϵ.

Proof. We first show that $0 < s < \infty$. Notice that since all communicating classes are singletons, $j \in \mathcal{T}$ iff $1 > \left(M_0\right)_{j,j}$. By Theorem 7.4 a), this is equivalent to

$$\begin{aligned} 0 \;<\; & 1 - \left(M_0\right)_{j,j} = \lim_{\epsilon \to 0^+} \left(1 - \left(M_\epsilon\right)_{j,j}\right) = \lim_{\epsilon \to 0^+} \sum_{i \neq j} \left(M_\epsilon\right)_{i,j} \\ = \; & \sum_{i \neq j} \lim_{\epsilon \to 0^+} \left(M_\epsilon\right)_{i,j} = \sum_{i \neq j} [R(M_\epsilon)_{i,j} = 0] C(M_\epsilon)_{i,j} \end{aligned}$$

By definition, s is the minimum resistance of the outgoing edges (excluding self-loops) of the closed classes of M_0, i.e., indices $j \notin \mathcal{T}$. For such j and all $i \neq j$, we must have $[R(M_\epsilon)_{i,j} = 0] C(M_\epsilon)_{i,j} = 0$. This implies that $R(M_\epsilon)_{i,j} > 0$ for $i \neq j$, so that $0 < s$.

By Lemma 1.5, each closed class/vertex of M_0 is contained in the closed class of M_ϵ. Thus, if M_0 possesses at least two closed classes, since both both vertices are in the closed class of M_ϵ, there are paths in $G_-\left(M_\epsilon\right)$ between them; in particular, there are outgoing edges in $G_-\left(M_\epsilon\right)$ from them, so that $s < \infty$. Alternatively, if M_0 possesses only one closed class/vertex, but M_ϵ is irreducible, there are paths in $G_-\left(M_\epsilon\right)$ from that vertex to every other vertex; in particular, there is at least one outgoing edge, so that $s < \infty$ in this case, as well.

Likewise, since s is a minimum, there exists some $i \neq j \notin \mathcal{T}$, such that $s = R\left(M_\epsilon\right)_{i,j}$. Since $s < \infty$, we must then have $C\left(M_\epsilon\right)_{i,j} > 0$, as well. Thus, $c = 2\max_{j\notin\mathcal{T}} \sum_{s=R(M_\epsilon)_{i,j}} C\left(M_\epsilon\right)_{i,j} > 0$.

Now observe that the entries of i_ϵ are either non-negative constants or $f(\epsilon)$, all of which are functions in $\mathcal{C}^+$, so that $i_\epsilon \in \mathrm{Pert}(n)$. Moreover, for $\epsilon \succ 0$, $f(\epsilon) > 0$, so that $0 < \left(i_\epsilon\right)_{j,j}$ for $j \in S_n$. We now show that, for $j \in S_n$ and $\epsilon \succ 0$, $\left(i_\epsilon\right)_{j,j}\left(1 - \left(M_\epsilon\right)_{j,j}\right) \leq 1$, so that, for $\epsilon \succ 0$, Lemma 5.3 implies that $\overline{M}_\epsilon \equiv \left(M_\epsilon\right)_{i_\epsilon}$ is Markov (by continuity, $\overline{M}_0$ is Markov, as well) and $\overline{M}_\epsilon \approx_{i_\epsilon} M_\epsilon$. Since $\left(M_\epsilon\right)_{j,j} \geq 0$, $\left(i_\epsilon\right)_{j,j}\left(1 - \left(M_\epsilon\right)_{j,j}\right) \leq \left(i_\epsilon\right)_{j,j}$. But either $\left(i_\epsilon\right)_{j,j} = 1$ or, since $R(f) = s > 0$, Theorem 7.4 a) implies that $\left(i_\epsilon\right)_{j,j} = f(\epsilon) \leq 1$ for $\epsilon \succ 0$.

Since, for $\epsilon \succeq 0$, we are only multiplying some off-diagonal entries by the non-zero number, $f(\epsilon)$, $G_-\left(\overline{M}_\epsilon\right)$ is the same as $G_-\left(M_\epsilon\right)$, except for possible differences is their self-loops. In particular, since M_ϵ is unichain, so is $\overline{M}_\epsilon$. More generally, since we are only multiplying some off-diagonal entries by $f \in \mathcal{C}^+$, $\left(\overline{M}_\epsilon\right)_{i,j} \in \mathcal{C}^+$ for $i \neq j$.

To show that $\overline{M}_\epsilon \in \mathrm{Pert}(n)$, since $\left(\overline{M}_\epsilon\right)_{j,j} = \left(M_\epsilon\right)_{j,j} \in \mathcal{C}^+$ for $j \notin \mathcal{T}$, it only remains to verify that $\left(\overline{M}_\epsilon\right)_{j,j} \in \mathcal{C}^+$ for $j \in \mathcal{T}$. In this case, since $\left(\overline{M}_\epsilon\right)_{j,j} = f(\epsilon)\left(M_\epsilon\right)_{j,j} + 1 - f(\epsilon)$, by parts b) and d) of Theorem 7.4, it suffices to show that $1 - f(\epsilon) \in \mathcal{C}^+$. Since $R(f) = s > 0 = R(1)$, we may apply Theorem 7.4 c) to conclude that $1 - f(\epsilon) \in \mathcal{C}^+$. Thus, $\overline{M}_\epsilon \in \mathrm{Pert}(n)$.

Next, we verify that f satisfies the assumptions of Lemma 7.11 with respect to $\overline{M}_\epsilon$. Starting with a), notice that, for $i \neq j$, $\left(\overline{M}_\epsilon\right)_{i,j}$ either equals $f(\epsilon)\left(M_\epsilon\right)_{i,j}$, if $j \in \mathcal{T}$, or $\left(M_\epsilon\right)_{i,j}$, if $j \notin \mathcal{T}$. In the former case, $R(f) \leq R\left(\overline{M}_\epsilon\right)_{i,j}$, by Theorem 7.4 d). In the latter, the inequality follows by construction, since $R(f) = s \leq R\left(\overline{M}_\epsilon\right)_{i,j}$ for all $i \neq j$ and $j \notin \mathcal{T}$. Thus, we have verified assumption a) of Theorem 7.11.

Now observe that, by Theorem 7.4 d), if $j \in \mathcal{T}$, $\left(\overline{M}_\epsilon\right)_{j,j} + f(\epsilon) - 1 = f(\epsilon)\left(M_\epsilon\right)_{j,j} \in \mathcal{C}^+$ and $R(f) \leq R\left(f(\epsilon)\left(M_\epsilon\right)_{j,j}\right)$. Thus, assumptions b) and c) are satisfied, in this case. On the other hand, if $j \notin \mathcal{T}$, $\left(\overline{M}_\epsilon\right)_{j,j} + f(\epsilon) - 1 = \left(M_\epsilon\right)_{j,j} + f(\epsilon) - 1 = f(\epsilon) - \left[\sum_{i\neq j}\left(M_\epsilon\right)_{i,j}\right]$.

We will again want to apply Theorem 7.4 c) to show that this is in $\mathcal{C}^+$, as well, with resistance no less than $R(f) = s$. There are two cases to consider. For a given $j \notin \mathcal{T}$, if there is some $i \neq j$ such that $R(f) = s = R\left(M_\epsilon\right)_{i,j}$, then $R\left(\sum_{i\neq j}\left(M_\epsilon\right)_{i,j}\right) = \min_{i\neq j} R\left(M_\epsilon\right)_{i,j} = s = R(f)$. In this case, $C\left(\sum_{i\neq j}\left(M_\epsilon\right)_{i,j}\right) = \sum_{i\neq j}[s = R\left(M_\epsilon\right)_{i,j}]C\left(M_\epsilon\right)_{i,j} = s = R(f) \leq \frac{c}{2} < c = C(f)$. Otherwise, $R(f) = s < R\left(M_\epsilon\right)_{i,j}$ for all $i \neq j$, so that

$R\left(\sum_{i\neq j}(M_\epsilon)_{i,j}\right) = \min_{i\neq j} R(M_\epsilon)_{i,j} > R(f)$. In either case, the conditions of Theorem 7.4 c) are satisfied, so that $f(\epsilon) - \left[\sum_{i\neq j}(M_\epsilon)_{i,j}\right] \in \mathcal{C}^+$ with resistance, $R(f)$. Thus, we have verified assumptions b) and b) of Theorem 7.11.

We now show that $\overline{M}_\epsilon$ is i_ϵ-equivalent to M_ϵ. Since we already know that $\overline{M}_\epsilon \approx_{i_\epsilon} M_\epsilon$ for $\epsilon \succ 0$, it only remains to show that $R\left(\left\|i_\epsilon \operatorname{stab}\left(M'_\epsilon\right)\right\|_1\right) = 0$. As in the proof of Theorem 7.15, we have $f(\epsilon)\operatorname{stab}(M_\epsilon) = i_\epsilon \operatorname{stab}\left(M'_\epsilon\right)$ for some $f \in \mathcal{C}^+$, specifically, $f(\epsilon) = Ji_\epsilon \operatorname{stab}\left(M'_\epsilon\right) = \left\|i_\epsilon \operatorname{stab}\left(M'_\epsilon\right)\right\|_1$. Thus, we must show that $R(f) = 0$.

Therefore, if $v_\epsilon = \operatorname{stab}(M_\epsilon)$ and $v'_\epsilon = \operatorname{stab}\left(M'_\epsilon\right)$, $i_\epsilon v'_\epsilon = f(\epsilon)v_\epsilon$, where $f(\epsilon) = \left\|i_\epsilon v'_\epsilon\right\|_1$. It remains to show that $r(f) = 0$. In particular, since $\left(i_\epsilon v'_\epsilon\right)_j = \left(v'_\epsilon\right)_j$ for $j \notin \mathcal{T}$, it suffices to show that $\operatorname{supp}_{v'_0} \cap \overline{\mathcal{T}} \neq \emptyset$. In this case, $\lim_{0^+} f(\epsilon) \geq \lim_{0^+}\left(i_\epsilon v'_\epsilon\right)_j = \lim_{0^+}\left(v'_\epsilon\right)_j > 0$, for $j \in \operatorname{supp}_{v'_0} \cap \overline{\mathcal{T}}$, so that $r(f) = 0$.

To see this, let $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ be the closed classes of $\mathcal{G}(M_0)$, so that $\overline{\mathcal{T}} = \bigcup_{j=1}^m \mathcal{C}_i$. Likewise, let $\left\{\mathcal{C}'_1, \ldots, \mathcal{C}'_{m'}\right\}$ be the closed classes of $\mathcal{G}\left(M'_0\right)$. By Theorem **??**, $\operatorname{supp}_{v'_0} = \bigcup_{j'\in\mathcal{J}'} \mathcal{C}'_{j'}$ for some $\mathcal{J}' \subset \left\{1, \ldots, m'\right\}$. Now observe that $\mathcal{G}(M_0) \subset \mathcal{G}\left(M'_0\right)$, so that, by Lemma 1.5, for every $1 \leq j' \leq m'$, there is $1 \leq j \leq m$ so that $\mathcal{C}'_{j'} \cap \mathcal{C}_j \neq \emptyset$. In particular, for any $j' \in \mathcal{J}$, $\emptyset \neq \mathcal{C}'_{j'} \cap \bigcup_{j=1}^m \mathcal{C}_i \subset \operatorname{supp}_{v'_0} \cap \overline{\mathcal{T}}$ □

By repeatedly applying Lemma 7.16, we may then guarantee that M_ϵ always possesses a non-trivial communicating class, as long as we keep track of the corresponding shift in virtual energies.

Corollary 7.17. *Given any $n \times n$ perturbed Markov matrix, M_ϵ, if M_0 possesses more than one closed class, there is an i_ϵ-equivalent perturbed Markov matrix, M'_ϵ, such that either M'_0 possesses a non-trivial communicating class or a single closed class, where $i_\epsilon \in \operatorname{Mat}_n\left(\mathcal{C}^+\right)$ is a diagonal matrix. In particular, if we denote the stable distributions of M_ϵ and M'_ϵ by v_ϵ and v'_ϵ, respectively, then $v_\epsilon = \frac{i_\epsilon v'_\epsilon}{\left\|i_\epsilon v'_\epsilon\right\|_1}$ and $r\left((v_\epsilon)_s\right) = r\left(\left(v'_\epsilon\right)_s\right) + r\left((i_\epsilon)_{s,s}\right)$. Similarly, if M_ϵ is irreducible with $n > 1$, there is such a matrix, M'_ϵ, where M'_0 possesses a non-trivial communicating class and M'_ϵ is irreducible.*

Proof. In any case, we will proceed by repeatedly applying Lemma 7.16. In other words, at each step, $M'_\epsilon = (M_\epsilon)_{D_\epsilon}$, for a diagonal matrix, D_ϵ, so that $\mathcal{P}\left(M'_\epsilon\right) = \mathcal{P}(M_\epsilon)$. In particular, if M_ϵ is irreducible, then so is M'_ϵ.

Now if M_0 possesses a non-trivial communicating class, then we may take $i_\epsilon = I$. Otherwise, if M_0 possesses more than one closed class or M_ϵ is irreducible with $n > 1$, we

may apply Lemma 7.16 to obtain an i_ϵ-equivalent perturbed Markov matrix, M'_ϵ, so that $\mathcal{G}\left(M'_0\right) \supset \mathcal{G}\left(M_0\right)$ is a strictly larger graph. We may repeat this construction until either M'_0 possesses a non-trivial communicating class or, if M_ϵ is not irreducible, M_0 possesses only one closed class. This is guaranteed to terminate by the time $\mathcal{G}\left(M'_0\right)$ is complete, if not sooner. □

7.5 Equivalence of PMMs and the MCCT

Our algorithm in section 7.7 will consist of alternately applying the scaling construction from Lemma 7.16 and the reduction construction of Theorem 5.6, generalized to PMMs. However, since reduction involves matrix inversion, we will want to choose the sets of indices to eliminate, so that we will only need to invert *real*-valued (i.e., zero-resistance) matrices, so that the required calculations are tractable. In this section, we appeal to the Markov Chain Tree Theorem to show how this is always possible. Specifically, we will show that

- given an $M_\epsilon \in \mathrm{PMM}(n)$, we may find an equivalent one for which the off-diagonal zero-resistance terms are actually constant;
- we may also find an equivalent one for which all edges within an SCC (excluding self-loops) are constant; and
- we may construct an equivalent one which, for each non-zero resistance path from v_i to v_j in $G\left(M_\epsilon\right)$, contains a edge from v_i to v_j with the same resistance.

We first show that the stochastically stable distribution of a perturbed Markov matrix depends only on the off-diagonal entries of the constant term, $C_0\left(M_\epsilon\right)$, of its cost matrix and the resistance matrix, $R\left(M_\epsilon\right)$.

Theorem 7.18. *Assume that M_ϵ and M'_ϵ are perturbed Markov matrices. If*

- $\overline{T}\left(M_\epsilon, i\right) = \overline{T}\left(M'_\epsilon, i\right)$,
- $R\left(M_\epsilon\right)_{\sigma(j),j} = R\left(M'_\epsilon\right)_{\sigma(j),j}, \forall j \neq i, \sigma \in \overline{T}\left(M_\epsilon, i\right)$

for every $1 \leq i \leq n$, then $M_\epsilon \backsim M'_\epsilon$. If, in addition,

- $C\left(M_\epsilon\right)_{\sigma(j),j} = C\left(M'_\epsilon\right)_{\sigma(j),j}, \forall j \neq \sigma\left(j\right), \sigma \in \bigcup_{i \in S_n} \overline{T}\left(M_\epsilon, i\right)$,

then $M_\epsilon \sim M'_\epsilon$.

Proof. Given M_ϵ and M'_ϵ satisfying the first three conditions in the statement of the Corollary, $r_i = \min_{\sigma \in T(M_\epsilon, i)} r\left(M_\epsilon, \sigma\right) = \min_{\sigma \in T\left(M'_\epsilon, i\right)} r\left(M_\epsilon, \sigma'\right)$, for $1 \leq i \leq n$. Thus, Theorem 7.7 implies that $r\left(\left(v_\epsilon\right)_i\right) = r\left(\left(v'_\epsilon\right)_i\right)$, so that $M_\epsilon \backsim M'_\epsilon$. Likewise, $\sum_{\sigma \in \overline{T}(M_\epsilon, i)} c\left(M_\epsilon, \sigma\right) = \sum_{\sigma \in \overline{T}\left(M'_\epsilon, i\right)} c\left(M_\epsilon, \sigma'\right)$, and $\sum_{\sigma \in \overline{T}(M_\epsilon)} c\left(M_\epsilon, \sigma\right) = \sum_{\sigma \in \overline{T}\left(M'_\epsilon\right)} c\left(M_\epsilon, \sigma'\right)$. Therefore, Lemma 7.7 implies that M_ϵ and M'_ϵ have identical stochastically stable distributions, so that $M_\epsilon \sim M'_\epsilon$. □

Corollary 7.19. *M_ϵ is equivalent to another perturbed Markov matrix, M'_ϵ, such that $C_\epsilon\left(M'_\epsilon\right)_{i,j}$ is constant with respect to ϵ, if $i \neq j$, and $R\left(M'_\epsilon\right)_{i,i} = 0$ for all i.*

Proof. Consider $\left(M_\epsilon\right)_{s,s} = \frac{1}{2}I + \frac{1}{2}M_\epsilon$, a matrix whose diagonal entries necessarily have 0-resistance. By Theorem 7.18, $M_\epsilon \sim \left(M_\epsilon\right)_{s,s}$. Further, denote the cost and resistance matrices of $\left(M_\epsilon\right)_{s,s}$ by $\overline{C}_\epsilon$ and $\overline{R}$, respectively, and define

$$\left(M'_\epsilon\right)_{i,j} = \begin{cases} 1 - \sum_{k \neq j} \epsilon^{\overline{R}_{k,j}} \left(\overline{C}_0\right)_{k,j} & \text{if } i = j \\ \epsilon^{\overline{R}_{i,j}} \left(\overline{C}_0\right)_{i,j} & \text{otherwise} \end{cases}$$

Notice that $\overline{R} = R\left(M'_\epsilon\right)$. This is clear for all off-diagonal entries. Moreover, they both have 0-resistance on the diagonal. In fact, $\left(M'_0\right)_{i,i} = \left(\overline{M}_0\right)_{i,i} = \frac{1+(M_0)_{i,i}}{2} > 0$. This also insures that $\left(M'_\epsilon\right)_{i,i} \succeq 0$, so that M'_ϵ is Markov. Moreover, $\overline{C}_0 = C_0\left(M'_\epsilon\right)$, while $\mathcal{G}\left(\left(M_\epsilon\right)_{s,s}\right)$ and $\mathcal{G}\left(M'_\epsilon\right)$ share the same underlying unweighted graph, $\mathcal{G}\left(\overline{R}\right)$. In particular, $\mathcal{P}\left(\left(M_\epsilon\right)_{s,s}\right) = \mathcal{P}\left(M'_\epsilon\right)$. Therefore, by Theorem 7.18, $M_\epsilon \sim \left(M_\epsilon\right)_{s,s} \sim M'_\epsilon$. □

Notice that this means that, if we are only interested in computing stochastically stable distributions, we may represent a perturbed Markov matrix simply by the two $n \times n$ matrices, $C_0\left(M_\epsilon\right)$ and $R\left(M_\epsilon\right)$, and we may assume that $r_{i,i} = 0, 1 \leq i \leq n$.

Example 7.20. We may briefly illustrate the construction of Corollary 7.19 as follows:

$$\begin{pmatrix} 0 & \cdots \\ 1-\epsilon & \\ \epsilon & \end{pmatrix} \rightarrow \begin{pmatrix} \frac{1}{2} & \cdots \\ \frac{1}{2} - \frac{\epsilon}{2} & \\ \frac{\epsilon}{2} & \end{pmatrix} \rightarrow \begin{pmatrix} \frac{1}{2} - \frac{\epsilon}{2} & \cdots \\ \frac{1}{2} & \\ \frac{\epsilon}{2} & \end{pmatrix}$$

□

As we have seen in Chapter 1, the self-loops of a Markov matrix may be adjusted to a certain degree without significantly affecting its dynamics. Similarly, we may specify a perturbed Markov matrix, with 0-resistance diagonal elements (such as in Corollary 7.19), by simply specifying its off-diagonal elements, so long as their column sums when $\epsilon = 0$ are all strictly less than 1.

Lemma 7.21. *If $M_\epsilon \in \mathrm{Mat}_n\left(\mathcal{C}^+\right)$ with $\sum_{i\neq j}\left(M_0\right)_{i,j} < 1$ for $1 \leq j \leq n$ and $\mathcal{P}\left(M_\epsilon\right)$ is starry for $\epsilon \succ 0$, then there is a unique[8] perturbed Markov matrix, M_ϵ, such that $\left(M'_\epsilon\right)_{i,j} = \left(M_\epsilon\right)_{i,j}$ for $1 \leq i \neq j \leq n$.*

Proof. For $1 \leq i \neq j \leq n$, let $\left(M'_\epsilon\right)_{i,j} = \left(M_\epsilon\right)_{i,j}$, and define $\left(M'_\epsilon\right)_{j,j} = 1 - \sum_{i\neq j}\left(M_\epsilon\right)_{i,j}$ for $1 \leq j \leq n$. Now observe that by continuity $\sum_{i\neq j}\left(M_\epsilon\right)_{i,j} \prec 1$, so that $\left(M'_\epsilon\right)_{j,j} \succ 0$. Therefore, $\left(M'_\epsilon\right)_{j,j} \in \mathcal{C}^+$ and $R\left(M'_\epsilon\right)_{j,j} = 0$. In particular, $R\left(M'_\epsilon\right) = R\left(M_\epsilon\right)$, so that $\mathcal{P}\left(M'_\epsilon\right)$ is starry and, by Lemma 6.2, M'_ϵ is regular for $\epsilon \succ 0$. Thus, M'_ϵ is a perturbed Markov matrix of the desired form. □

Given a regular, perturbed Markov matrix, M_ϵ, consider a submatrix of M_ϵ consisting of all the transitions within a given communicating class of M_0. The following Lemma shows that, without loss of generality, we may assume that such a submatrix is constant off the diagonal. For instance, consider Example 7.8: we may drop the $(3,2)$– and $(2,3)$–entries (adjusting the diagonal entries accordingly) without changing its stochastically stable distribution, since they are not in any of the minimal resistance spanning trees.

Lemma 7.22.

For any regular, perturbed Markov matrix, M_ϵ, there is an equivalent one, M'_ϵ such that $M'_0 = M_0$ and if $[p,q] \in \mathcal{P}\left(M'_0\right)$, $R\left(M'_\epsilon\right)_{p,q} = 0$ or ∞.

Proof. By Corollary 7.19, we may assume that $r_{i,i} = 0$, $1 \leq i \leq n$. Consider any arbitrary pair of distinct states, p and q, in the same communicating class of M_0 with a non-zero resistance edge, $r_{p,q} > 0$, between them. Notice that this implies $p \neq q$. In this case, we may define $M'_\epsilon = M_\epsilon - \left(M_\epsilon\right)_{p,q}\left(e_{p,q} - e_{q,q}\right)$, where $\left\{e_{i,j}\right\}$ is the standard basis for $\mathrm{Mat}_n\left(\mathbb{R}\right)$. This is just M_ϵ without an edge from q to p. Since $[p,q] \in \mathcal{P}\left(M_0\right)$, and $M'_0 = M_0$, $[p,q] \in \mathcal{P}\left(M'_0\right)$, and $\mathcal{P}\left(M'_\epsilon\right) = \mathcal{P}\left(M_\epsilon\right)$. In particular, M'_ϵ is regular. Since

[8] Up to equality in $\mathrm{Mat}_n\left(\mathcal{C}\right)$.

$\left(M'_\epsilon\right)_{p,q} = 0$ and $\left(M'_\epsilon\right)_{q,q} = (M_\epsilon)_{p,q} + (M_\epsilon)_{q,q} \succeq 0$, M'_ϵ is Markov for each ϵ and, by Theorem 7.4, all entries have positive resistance. Thus, M'_ϵ is a perturbed Markov matrix.

By Corollary 7.19, to show that $M'_\epsilon \sim M_\epsilon$, it suffices to verify that the edge from q to p is not part of any minimal spanning tree of M_ϵ. Specifically, if $\sigma \in T(M_\epsilon, k)$ is a spanning subtree rooted at k containing this edge (i.e., $\sigma(q) = p$), then the resistance of σ is not minimal (i.e., $\sigma \notin \overline{T}(M_\epsilon, k)$). Therefore, assume we are given $\sigma \in T(M_\epsilon, k)$ with $\sigma(q) = p$ and consider the associated directed spanning tree, $\mathcal{T}$.

Let $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ be the strongly connected components of $\mathcal{G}(M_0)$. As in Chapter 6, $\mathcal{T}$ defines a pre-order on the vertices of $\mathcal{G}(M_\epsilon)$ so that $\sigma(j) \prec j$ for $j \neq k$. By re-indexing, if necessary, we may assume that the states of M_ϵ have been indexed to respect this preorder, so that $i \prec j \Rightarrow i < j$, i.e., ancestors always have smaller indices. In particular, we may assume that $k = 1$. This induces a preorder on the $\mathcal{C}_s$ so that $\mathcal{C}_s \prec \mathcal{C}_t$ iff $\min \mathcal{C}_s < \min \mathcal{C}_t$. By re-indexing again, we may assume that $\min \mathcal{C}_s < \min \mathcal{C}_t \Rightarrow s < t$. In particular, we must have $1 \in \mathcal{C}_1$.

We may now proceed by induction to construct a new directed spanning tree, $\overline{\mathcal{T}}$ corresponding to $\overline{\sigma}$, with strictly smaller resistance than $\mathcal{T}$. Specifically, we will construct a spanning tree rooted at 1 which only contains 0-resistance edges within each $\mathcal{C}_s$. Moreover, there will be at most one edge in $\overline{\mathcal{T}}$ between distinct communicating classes, i.e., $\forall s > 1, \exists j_s \in \mathcal{C}_s \text{ s.t. } \overline{\sigma}(j_s) = \sigma(j_s) \in \mathcal{C}_u$ and $u < s$. Intuitively, we will choose edges of $\mathcal{T}$ to build a directed tree on the communicating classes (specified by j_s), which will serve to link a set of 0-resistance spanning trees spanning each class into an directed tree rooted at 1.

Since $\mathcal{C}_1$ is a strongly connected component of $\mathcal{G}(M_0)$, by Lemma 6.2, it contains an directed tree rooted at 1 consisting entirely of 0-resistance edges spanning $\mathcal{G}(M_0)|_{\mathcal{C}_1}$, which defines $\overline{\sigma}(j)$ for $j \in \mathcal{C}_1$. For any $1 < t \leq m$, assume that we have defined $\overline{\sigma}$ over $\bigcup_{s=1}^{t-1} \mathcal{C}_s$ and a sequence $j_s \in \mathcal{C}_s$ such that $\forall 1 < s < t, \overline{\sigma}(j_s) = \sigma(j_s) \in \mathcal{C}_u$ for some $u < s$. It suffices to show that we may extend $\overline{\sigma}$ to $\mathcal{C}_t$ and define $j_t \in \mathcal{C}_t$ so that $\overline{\sigma}(j_t) = \sigma(j_t) \in \mathcal{C}_u$ for some $u < t$.

Let $j_t = \min \mathcal{C}_t$. Since $\sigma(j_t) < t$, if $\sigma(j_t) \in \mathcal{C}_u$, $\min \mathcal{C}_u < j_t = \min \mathcal{C}_t$, so that $u < t$. As before, Lemma 6.2 guarantees the existence of an directed tree rooted at j_t spanning $\mathcal{G}(M_0)|_{\mathcal{C}_t}$. This defines $\overline{\sigma}$ on $\mathcal{C}_t - \{j_t\}$ and we take $\overline{\sigma}(j_t) = \sigma(j_t)$. By induction, we eventually obtain a subgraph of $\mathcal{G}(M_\epsilon)$ consisting of $n-1$ edges which contains a path

from each vertex of $\mathcal{G}(M_\epsilon)$ to 1, which is thus an directed spanning tree rooted at 1.

Since $\overline{\mathcal{T}}$ only contains edges of $\mathcal{T}$, plus edges with resistance 0, and we know it does not include the edge from q to p, $r(\overline{\sigma}, M_\epsilon) \leq r(M_\epsilon, \sigma) - r_{p,q} < r(M_\epsilon, \sigma)$. In particular, σ did not have minimal resistance, the edge from q to p is not part of any minimal spanning tree of M_ϵ, and $T(M_\epsilon, k) = T\left(M'_\epsilon, k\right)$ for arbitrary k.

Since all other off-diagonal entries of M_ϵ and M'_ϵ agree, all necessary entries of the resistance and cost matrices agree. In particular, Corollary 7.19 implies that $M'_\epsilon \sim M_\epsilon$. By repeating this construction sufficiently many times, we may eliminate all non-zero resistance edges within any communicating class. □

Lemma 7.23. *If* $0 < r_{i,j}, r_{j,k} < \infty$, $f \succeq 0$, *and* $r(f) = r_{i,j} + r_{j,k}$, *then* $M_\epsilon \simeq M'_\epsilon = M_\epsilon + f(\epsilon)\left(e_{i,k} - e_{k,k}\right)$.

Proof. As in the proof of Lemma 7.22, M'_ϵ is a Markov matrix for each ϵ with entries in $\mathcal{C}^+$. If $(M_\epsilon)_{i,k} \not\approx 0$, $\mathcal{G}\left(M'_\epsilon\right) = \mathcal{G}(M_\epsilon)$. Otherwise, since $r_{i,j}, r_{j,k} < \infty$, $(M_\epsilon)_{i,j}, (M_\epsilon)_{j,k} \not\approx 0$, so that there is an edge from e_k to e_i in $\mathcal{G}(M_\epsilon)$. In either case, $\mathcal{P}\left(M'_\epsilon\right) = \mathcal{P}(M_\epsilon)$, so that M'_ϵ is a regular, perturbed Markov matrix. It remains to show that $M_\epsilon \simeq M'_\epsilon$.

Observe that $\left(M'_\epsilon\right)_{i,k} = (M_\epsilon)_{i,k} + f(\epsilon)$. If $r_{i,k} < r_{i,j} + r_{j,k}$, Theorem 7.4 implies that $r\left(\left(M'_\epsilon\right)_{i,k}\right) = r_{i,k}$. In this case, Corollary 7.19 implies that $M_\epsilon \simeq M'_\epsilon$.

Now consider the case when $r_{i,k} = r_{i,j} + r_{j,k}$. We will show that if $\sigma \in T(M_\epsilon, s)$ is a spanning subtree rooted at s with $\sigma(k) = i$, then there is another spanning subtree, $\sigma \in T(M_\epsilon, s)$, rooted at s with $\sigma'(k) = j$, and $r\left(\sigma', M_\epsilon\right) \leq r(M_\epsilon, \sigma)$. Thus, by Theorem 7.18 $M_\epsilon \simeq M'_\epsilon$. Therefore, assume we are given $\sigma \in T(M_\epsilon, s)$ with $\sigma(k) = i$ and consider the associated directed spanning tree, $\mathcal{T}$. If we remove the edge (e_k, e_i), we are left with two directed subtrees, $\mathcal{T}_1$ and $\mathcal{T}_2$, where we may assume that i and s are in $\mathcal{T}_1$ and k is in $\mathcal{T}_2$. Node j must be in one of these subtrees. If j is in $\mathcal{T}_1$, by adding the edge from k to j we obtain a tree with total resistance which has decreased by $r_{i,j}$. Formally, defining $\sigma'(t) = \sigma(t)$, for all $t \neq k$, and $\sigma'(k) = j$, $r\left(\sigma', M_\epsilon\right) = r(M_\epsilon, \sigma) - r_{i,j} < r(M_\epsilon, \sigma)$.

Otherwise, j is in $\mathcal{T}_2$. Deleting the edge from j in $\mathcal{T}_2$ splits it into two smaller trees, $\mathcal{T}'_2$ and $\mathcal{T}''_2$, where we may assume that j is in $\mathcal{T}''_2$ and k is in $\mathcal{T}'_2$. By adding the edges from k to j and from j to i, we obtain a tree with total resistance no greater than before, but which does not include the edge from k to i. Formally, defining $\sigma'(t) = \sigma(t)$, for all $t \neq k, j$, $\sigma'(k) = j$, $\sigma'(j) = i$, $r\left(\sigma', M_\epsilon\right) \leq r(M_\epsilon, \sigma) - r_{i,k} + r_{i,j} + r_{j,k} = r(M_\epsilon, \sigma)$. □

7.6 Reduction of PMMs

We now wish to generalize the reduction construction of Chapter 5 so that we may apply it to a perturbed Markov matrix. Since the definition of the reduction, given in Theorem 5.6, was originally stated in terms of matrix inverses and subtraction (which are problematic in the class of perturbed matrices), it is not obvious that the result is a perturbed Markov matrix. As with scaling, this will require careful analysis. As mentioned in section 7.5, our algorithm in section 7.7 is guided by the need to eliminate sets of indices which only require the inversion of *real*-valued (i.e., zero-resistance) matrices. Thus, our calculations will be guided by the zero-resistance subgraph, $G_{-}(M_0)$ of $G_{-}(M_\epsilon)$.

Specifically, in this section we will:

- show that if $s \subset S_n$ is open with respect to M_0, it is also open with respect to $M_\epsilon \in \mathrm{PMM}(n)$;
- generalize Theorem 4.5 to PMMs, proving that $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathrm{Pert}(|s|)$ for $M_\epsilon \in \mathrm{PMM}(n)$ and $s \subset S_n$ open, giving formulas for its resistance and cost matrices in terms of minimum resistance walks (cf. Theorems 5.8 and 7.7);
- show that if we apply the reduction construction to an $M_\epsilon \in \mathrm{PMM}(n)$ for $\epsilon \succ 0$, with respect to a fixed open set of indices, s, we obtain an $\widehat{M_\epsilon} \in \mathrm{PMM}(|\overline{s}|)$;
- we derive formulas for the resistance and cost matrices of the reduction in terms of the resistance and cost matrices of M_ϵ, which shows that reduction preserves asymptotic equality;
- generalize Theorem 5.12 to PMMs;
- show that reduction preserves equivalence of PMMs defined in section 7.4; and
- show that for the purposes of computing $\mathrm{ssd}(M_\epsilon)$, we may compute the reduction while only inverting a constant matrix.

As mentioned above, we will want to apply the reduction construction to sets of indices which are open with respect to the (unweighted) zero-resistance subgraph, $G_{-}(M_0)$. We must first verify that this is feasible.

Lemma 7.24. *Let M_ϵ be an $n \times n$ perturbed Markov matrix. If $s \subset S_n$ is open with respect to M_0, it is open with respect to M_ϵ.*

Proof. By Lemma 4.5 b), since $s \subset S_n$ is open with respect to M_0, $I - (M_0)_{s,s} = \pi_s (I - M_0) \imath_s$ is invertible. Thus, $|\pi_s (I - M_0) \imath_s| \neq 0$, which implies that $|\pi_s (I - M_\epsilon) \imath_s| \neq 0$ for $\epsilon \succeq 0$. Thus, by the contrapositive of Lemma 4.5 c), $s \subset S_n$ must be open with respect to M_ϵ for $\epsilon \succ 0$, or equivalently, with respect to M_ϵ as a perturbed matrix. □

Lemma 7.24 implies that, for $\epsilon \succeq 0$, we may apply our reduction construction to any perturbed Markov matrix, M_ϵ, to eliminate any set of states, s, which are open with respect to the unperturbed matrix, M_0. We now show that the result is a perturbed Markov matrix.

The difficulty is that, in general, we cannot invert a perturbed matrix, since this might involve subtraction or division. However, in this specific case, we may express the inverse in terms of multiplication and addition alone. To do so, we will need to generalize some more notation, this time from section 5.2.

For any $\sigma \in \mathcal{S}_n(k)$, corresponding to a walk of length k in $G(M_\epsilon)$, we will define its weight in $G(M_\epsilon)$, $W(M_\epsilon, \sigma) = \Pi_{t \in S_k} \left((M_\epsilon)_{\sigma_t, \sigma_{t-1}} \right)$. Notice that this is simply (the equivalence class of) the function, which at $\epsilon \succeq 0$ is given by the weight of σ in M_ϵ.

By Corollary 7.5 b), $W(M_\epsilon, \sigma) \in \mathcal{C}^+$. Thus, we may also define the *resistance of σ in M_ϵ* as $R(M_\epsilon, \sigma) \equiv R(W(M_\epsilon, \sigma))$. Similarly, we define the *cost of σ in M_ϵ* as $C(M_\epsilon, \sigma) \equiv C(W(M_\epsilon, \sigma))$. By Corollary 7.5 b), the resistance and cost of the walk, σ, satisfy the following equations:

$$R(M_\epsilon, \sigma) = R(W(M_\epsilon, \sigma)) = R\left(\Pi_{t \in S_k} \left((M_\epsilon)_{\sigma_t, \sigma_{t-1}} \right) \right) = \sum_{t \in S_k} R\left((M_\epsilon)_{\sigma_t, \sigma_{t-1}} \right) \quad (7.3)$$

$$C(M_\epsilon, \sigma) = C(W(M_\epsilon, \sigma)) = C\left(\Pi_{t \in S_k} \left((M_\epsilon)_{\sigma_t, \sigma_{t-1}} \right) \right) = \Pi_{t \in S_k} C\left((M_\epsilon)_{\sigma_t, \sigma_{t-1}} \right) \quad (7.4)$$

Notice that although this notation exactly mirrors the case when σ represents a spanning tree, the meaning will be clear from the context, depending whether σ represents a tree or a walk.

Just as the stochastically stable distribution is defined in terms of minimum resistance spanning trees, the reduction is defined in terms of minimum resistance *walks*. However, since the collection of walks is infinite, we must argue that such a collection of minimum resistance walks is well-defined.

Lemma 7.25. *For any $n \times n$ perturbed Markov matrix, M_ϵ, set of indices $s \subset S_n$, and individual indices i and j, there exists a set $\mathcal{S}(M_\epsilon, s, i, j) \subset \mathcal{S}_n(s, i, j)$ such that, if $\sigma_1 \in \mathcal{S}(M_\epsilon, s, i, j)$ and $\sigma_2 \in \mathcal{S}_n(s, i, j)$, then $R(M_\epsilon, \sigma_1) \leq R(M_\epsilon, \sigma_2)$, i.e., the set of minimal resistance walks from j to i passing through s is well-defined. Moreover, $\min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma)$ exists and is equal to $R(M_\epsilon, \sigma)$ for all $\sigma \in \mathcal{S}(M_\epsilon, s, i, j)$.*

Proof. Consider the sets

$$\mathcal{R} \equiv \{R(M_\epsilon, \sigma) \mid \sigma \in \mathcal{S}_n(s, i, j)\} \text{ and } \overline{\mathcal{R}} \equiv \{R(M_\epsilon, \sigma) \mid \sigma \in \mathcal{S}_n(s, i, j) \text{ s. t. } \sigma \text{ is a path}\}$$

These are both positive sets of real numbers, so that $\inf \overline{\mathcal{R}}$ and $\inf \mathcal{R}$ are well-defined. Since any path is a walk, $\inf \overline{\mathcal{R}} \geq \inf \mathcal{R}$. Since we may convert any walk into a path, by simply dropping cycles from the walk, for each $r \in \mathcal{R}$, there is an $\overline{r} \in \overline{\mathcal{R}}$ such that $\overline{r} \leq r$. Thus, $\inf \overline{\mathcal{R}} \leq \inf \mathcal{R}$, so that $\inf \overline{\mathcal{R}} = \inf \mathcal{R}$. Finally, since there are only finitely many possible paths from j to i, $\inf \overline{\mathcal{R}} = \min \overline{\mathcal{R}}$. That is, the infimum is attained at some specific path. Since a path is a walk, and $\inf \overline{\mathcal{R}} = \inf \mathcal{R}$, the infimum is attained for $\mathcal{R}$, as well, so that $\min \overline{\mathcal{R}} = \inf \overline{\mathcal{R}} = \inf \mathcal{R} = \min \mathcal{R} = \min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma)$.

Define

$$\mathcal{S}(M_\epsilon, s, i, j) = \left\{ \sigma \in \mathcal{S}_n(s, i, j) \mid R(M_\epsilon, \sigma) = \min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma) \right\}$$

If $\sigma_1 \in \mathcal{S}(M_\epsilon, s, i, j)$ and $\sigma_2 \in \mathcal{S}_n(s, i, j)$, then $R(M_\epsilon, \sigma_1) = \min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma) \leq R(M_\epsilon, \sigma_2)$. By definition, $\min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma) = R(M_\epsilon, \sigma)$ for all $\sigma \in \mathcal{S}(M_\epsilon, s, i, j)$. □

We will partition this set of minimum resistance walks by their length, to define

$$\mathcal{S}(M_\epsilon, s, i, j, k) = \mathcal{S}_n(s, i, j, k) \cap \mathcal{S}(M_\epsilon, s, i, j)$$

as well.

We now show that $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathcal{C}^+$ and give formulas for its resistance and cost matrices in terms of minimum resistance walks.

Lemma 7.26. *Let M_ϵ be an $n \times n$ perturbed Markov matrix, $s \subset S_n$ be an open set of indices with respect to M_0, and $(M_\epsilon)_{s,s}$ the corresponding principal submatrix. Then $(M_\epsilon)_{s,s}^{(u)} = \sum_{j=0}^{u-1} (M_\epsilon)_{s,s}^j$ converges uniformly to $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathrm{Pert}(n, n)$. Moreover,*

$$\begin{aligned} R\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)_{i,j} &= \min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma) \text{ and} \\ C\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)_{i,j} &= \sum_{\sigma \in \mathcal{S}(M_\epsilon, s, i, j)} C(M_\epsilon, \sigma) \end{aligned}$$

Proof. Since M_ϵ has only finitely many entries, there is an interval, $[0, \delta]$, over which all of its entries, and hence all of the entries of $(M_\epsilon)^{(u)}_{s,s} = \sum_{j=0}^{u-1} (M_\epsilon)^j_{s,s}$, are continuous and non-negative. By assumption, s is open with respect to M_0, and by Lemma 7.24 with respect to M_ϵ for sufficiently small ϵ. Assume that δ has been chosen so that s is open with respect to M_ϵ for $\epsilon \in [0, \delta]$. Thus, by Lemma 4.5, $(M_\epsilon)^{(u)}_{s,s}$ converges to $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \geq 0$ for each $\epsilon \in [0, \delta]$.

Now observe that $\left\|(M_\epsilon)^n_{s,s}\right\|_1$ is a non-negative continuous function for $\epsilon \succeq 0$, since it is the maximum of a finite set of non-negative, continuous functions (i.e., the column sums of $(M_\epsilon)^n_{s,s}$). Since $\left\|(M_\epsilon)^n_{s,s}\right\|_1$ is bounded by some $0 \leq c < 1$ for each $\epsilon \in [0, \delta]$, we may assume that it is uniformly bounded on $[0, \delta]$ by $0 \leq c < 1$. In particular, $M^{(s)}_\epsilon$ converges uniformly to $\left(I - (M_\epsilon)_{s,s}\right)^{-1}$ on $[0, \delta]$.

Since $[0, \delta]$ is compact, the entries of $(M_\epsilon)^{(u)}_{s,s}$ are uniformly continuous on $[0, \delta]$ (Wheeden and Zygmund, 1977). Since $(M_\epsilon)^{(u)}_{s,s}$ converges uniformly to $\left(I - (M_\epsilon)_{s,s}\right)^{-1}$, the entries of $\left(I - (M_\epsilon)_{s,s}\right)^{-1}$ are continuous and non-negative on $[0, \delta]$ (Wheeden and Zygmund, 1977). Finally, if an entry of $\left(I - (M_\epsilon)_{s,s}\right)^{-1}$ is 0 for some $\epsilon \in (0, \delta]$, since this is the sum of a non-negative series, all the terms in the series must be 0. But all the terms are in $C^+[0, *]$, so that they must be identically 0, so that the limit must be in $C^+[0, *]$.

To show that $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathcal{C}^+$, we must compute the resistance and cost of each of its entries. Consider the $(i, j)^{\text{th}}$ entry of $M^{(k)}_\epsilon$, for some fixed $i, j \in S_n$. By Theorem 7.4,

$$e_i^t \, (M_\epsilon)^{(k)}_{s,s} \, e_j = \sum_{u=0}^{k-1} e_i^t \, (M_\epsilon)^u_{s,s} \, e_j = \sum_{t=0}^{k-1} \sum_{s \in \mathcal{P}(\mathcal{S},i,j,t)} (M_\epsilon)_s \in \mathcal{C}^+$$

Therefore, by Theorem 7.4 again,

$$r_k = R\left(M^{(k)}_\epsilon\right)_{i,j} = r\left(\sum_{t=0}^{k-1} \sum_{s \in \mathcal{P}(\mathcal{S},i,j,t)} (M_\epsilon)_s\right) = \min_{0 \leq t \leq k-1} \min_{s \in \mathcal{P}(\mathcal{S},i,j,t)} r\left(s, M_\epsilon\right)$$

Also observe that

$$\begin{aligned} \inf_k r_k &= \inf_k \min_{s \in \mathcal{P}(\mathcal{S},i,j,k)} r\left(s, M_\epsilon\right) \\ &= \inf_{s \in \mathcal{P}(\mathcal{S},i,j)} r\left(s, M_\epsilon\right) = \min_{s \in \mathcal{P}(\mathcal{S},i,j)} r\left(s, M_\epsilon\right) \end{aligned}$$

Now take

$$\begin{aligned} k' &= \min\left\{k \mid \overline{\mathcal{P}}\left(\mathcal{S},i,j\right)\cap\mathcal{P}\left(\mathcal{S},i,j,k\right)\neq\emptyset\right\} \text{ and} \\ s' &\in \overline{\mathcal{P}}\left(\mathcal{S},i,j\right)\cap\mathcal{P}\left(\mathcal{S},i,j,k'\right) \end{aligned}$$

This then implies that

$$r_{k'+1} = r\left(s', M_\epsilon\right) = \min_{s\in\mathcal{P}(\mathcal{S},i,j)} r\left(s, M_\epsilon\right) = \inf_k r_k$$

i.e., the infimum is attained for $k = k' + 1$, so it is a minimum. In particular, $r = \min_k r_k$ is well-defined and $R\left(M_\epsilon^{(k)}\right)_{i,j} = r$ for at least one value of k.

Now consider $\epsilon^{-r}\left(M_\epsilon^{(k)}\right)_{i,j} \in \mathcal{C}^+$, by Theorem 7.4.

$$\begin{aligned} \epsilon^{-r}\left(M_\epsilon^{(k)}\right)_{i,j} &= \epsilon^{-r}\sum_{t=0}^{k-1}\sum_{s\in\mathcal{P}(\mathcal{S},i,j,t)}\left(M_\epsilon\right)_s \\ &= \sum_{t=0}^{k-1}\sum_{s\in\mathcal{P}(\mathcal{S},i,j,t)}\epsilon^{-r}\left(M_\epsilon\right)_s \end{aligned}$$

Defining $\overline{\mathcal{P}}\left(\mathcal{S},i,j,k\right) = \overline{\mathcal{P}}\left(\mathcal{S},i,j\right)\cap\bigcup_{t=0}^{k-1}\mathcal{P}\left(\mathcal{S},i,j,t\right)$, we also have

$$\begin{aligned} c_k &= \lim_{0^+}\epsilon^{-r}\left(M_\epsilon^{(k)}\right)_{i,j} \\ &= \sum_{s\in\overline{\mathcal{P}}(\mathcal{S},i,j,k)} c_0\left(s, M_\epsilon\right) \end{aligned}$$

As before, $\epsilon^{-r}\left(M_\epsilon^{(k)}\right)_{i,j}$ is continuous and the limit, $\epsilon^{-r}\left(I - M_\epsilon\right)^{-1}_{i,j}$, is continuous on $[0,\delta]$. Therefore,

$$\lim_{0^+}\epsilon^{-r}\left(I - M_\epsilon\right)^{-1}_{i,j} = \lim_{k\to\infty} c_k = \sum_{s\in\overline{\mathcal{P}}(\mathcal{S},i,j)} c_0\left(s, M_\epsilon\right)$$

Moreover, c_k is a non-decreasing function of k, with $c_{k'+1} > 0$, so that $\lim_{0^+}\epsilon^{-r}\left(I - M_\epsilon\right)^{-1}_{i,j} > 0$. This implies that $r\left(\left(I - M_\epsilon\right)^{-1}_{i,j}\right) = r$ and $\left(I - \left(M_\epsilon\right)_{s,s}\right)^{-1} \in \mathrm{Mat}_n\left(\mathcal{C}^+\right)$. In addition, this shows that

$$C_0\left(\left(I - M_\epsilon\right)^{-1}\right)_{i,j} = \sum_{s\in\overline{\mathcal{P}}(\mathcal{S},i,j)} c_0\left(s, M_\epsilon\right)$$

□

We may now show that the reduction of a perturbed Markov matrix is also a perturbed Markov matrix. Moreover, the resistance of the (i,j)-th entry is minimum resistance of paths from j to i whose interior states are all in $\mathcal{S}$.

Theorem 7.27. *Under the assumptions of Lemma 7.26, for $\epsilon \succeq 0$, the reduction, $\widehat{M}_\epsilon$, with respect to $\mathcal{S}$ of M_ϵ is a perturbed Markov matrix with*

$$\begin{aligned} R\left(\widehat{M}_\epsilon\right)_{i,j} &= \min_{s\in\mathcal{P}(M,\mathcal{S},i,j)} r\left(s, M_\epsilon\right) \text{ and} \\ C_0\left(\widehat{M}_\epsilon\right)_{i,j} &= \sum_{s\in\overline{\mathcal{P}}(\mathcal{S},i,j)} c_0\left(s, M_\epsilon\right) \end{aligned}$$

Proof. Partitioning M_ϵ according to $\mathcal{S}$ gives $M_\epsilon = \begin{pmatrix} \widetilde{M}_\epsilon & \overline{N}_\epsilon \\ \widetilde{N}_\epsilon & \left(M_\epsilon\right)_{s,s} \end{pmatrix}$. If $|\mathcal{S}| = k$, $\widetilde{M}_\epsilon \in \mathrm{Mat}_{n-k}\left(\mathcal{C}^+\right)$, $\widetilde{N}_\epsilon \in \mathrm{Mat}_{k,n-k}\left(\mathcal{C}^+\right)$, and $\overline{N}_\epsilon \in \mathrm{Mat}_{n-k,k}\left(\mathcal{C}^+\right)$. By Lemma 7.26, $\left(I - \left(M_\epsilon\right)_{s,s}\right)^{-1} \in \mathrm{Mat}_k\left(\mathcal{C}^+\right)$. Since $\widehat{M}_\epsilon = \widetilde{M}_\epsilon + \overline{N}_\epsilon\left(I - \left(M_\epsilon\right)_{s,s}\right)^{-1}\widetilde{N}_\epsilon$, Theorem 7.4 implies that $\widehat{M}_\epsilon \in \mathrm{Mat}_{n-k}\left(\mathcal{C}^+\right)$. In addition, Theorem **??** guarantees that $\widehat{M}_\epsilon$ is regular. The formulas for its resistance and cost matrices follow directly from Lemmas 7.4 and 7.26. □

It is important to point out that our reduction construction preserves both equivalence and weak equivalence of perturbed Markov matrices. Also, Theorem 5.12 generalizes to this setting.

Theorem 7.28. *Assume that M_ϵ and M'_ϵ are perturbed Markov matrices, $\mathcal{S}$ is a set of states that does not include a closed class of M_0 or M'_0, and $\left(\widehat{M}_\epsilon, p_\epsilon, i_\epsilon\right)$ and $\left(\widehat{M}'_\epsilon, p'_\epsilon, i'_\epsilon\right)$ are the reductions of M_ϵ and M'_ϵ, respectively, with respect to $\mathcal{S}$. If $M_\epsilon \backsim M'_\epsilon$, then $\widehat{M}_\epsilon \backsim \widehat{M}'_\epsilon$. Likewise, if $M_\epsilon \sim M'_\epsilon$, then $\widehat{M}_\epsilon \sim \widehat{M}'_\epsilon$. Finally, $M_\epsilon \sim_{i_\epsilon} \widehat{M}_\epsilon$.*

Proof. By Theorem **??**, $\mathcal{P}\left(\widehat{M}_\epsilon\right) = \mathcal{P}\left(\widehat{M}'_\epsilon\right)$. Let v_ϵ and v'_ϵ be the stable distributions of M_ϵ and M'_ϵ, while $\widehat{v}_\epsilon$ and $\widehat{v}'_\epsilon$ are the stable distributions of $\widehat{M}_\epsilon$ and $\widehat{M}'_\epsilon$, respectively. Notice that, by Corollary **??**, $i_\epsilon\widehat{v}_\epsilon$ is an extension of $\widehat{v}_\epsilon$, which necessarily has 0-resistance entries. Thus, $i_\epsilon\widehat{v}_\epsilon$ must contain 0-resistance entries. In particular, by Theorem 7.4, $r\left(\left\|i_\epsilon\widehat{v}_\epsilon\right\|_1\right) = 0$, and likewise $r\left(\left\|i'_\epsilon\widehat{v}'_\epsilon\right\|_1\right) = 0$.

By Theorem 5.12, $\left\|i_\epsilon\widehat{v}_\epsilon\right\|_1 v_\epsilon = i_\epsilon\widehat{v}_\epsilon$ and $\left\|i'_\epsilon\widehat{v}'_\epsilon\right\|_1 v'_\epsilon = i'_\epsilon\widehat{v}'_\epsilon$ for $\epsilon \succeq 0$. If $M_\epsilon \backsim M'_\epsilon$, for every $s \in \overline{\mathcal{S}}$, $r\left(\left(\widehat{v}_\epsilon\right)_s\right) = r\left(\left(i_\epsilon\widehat{v}_\epsilon\right)_s\right) = r\left(\left\|i_\epsilon\widehat{v}_\epsilon\right\|_1\left(v_\epsilon\right)_s\right) = r\left(\left(v_\epsilon\right)_s\right) = r\left(\left(v'_\epsilon\right)_s\right) = r\left(\left\|i'_\epsilon\widehat{v}'_\epsilon\right\|_1\left(v'_\epsilon\right)_s\right) = r\left(\left(i'_\epsilon\widehat{v}'_\epsilon\right)_s\right) = r\left(\left(\widehat{v}'_\epsilon\right)_s\right)$. Thus, $\widehat{M}_\epsilon \backsim \widehat{M}'_\epsilon$.

If $M_\epsilon \sim M'_\epsilon$, $\frac{i_0\widehat{v}_0}{\|i_0\widehat{v}_0\|_1} = v_0 = v'_0 = \frac{i'_0\widehat{v}'_0}{\left\|i'_0\widehat{v}'_0\right\|_1}$. In particular, applying the common left-inverse, π, of section 5.2 to both sides $\frac{\widehat{v}_0}{\|i_0\widehat{v}_0\|_1} = \frac{\widehat{v}'_0}{\left\|i'_0\widehat{v}'_0\right\|_1}$. Since these are both distributions, we must in fact have $\widehat{v}_0 = \widehat{v}'_0$, so that $\widehat{M}_\epsilon \sim \widehat{M}'_\epsilon$.

Finally, notice how Theorem 5.12 implies that $M_\epsilon \smile_{i_\epsilon} \widehat{M}_\epsilon$. Lemma 7.26 insures that $(i_\epsilon)_{i,j} \in \mathcal{C}^+$. Finally, we have already observed that $r\left(\|i_\epsilon\widehat{v}_\epsilon\|_1\right) = 0$, so that $M_\epsilon \sim_{i_\epsilon} \widehat{M}_\epsilon$. □

We now show that we may compute the reduction while only inverting a constant matrix. For the following Lemma, let M_ϵ be a perturbed Markov matrix with a non-trivial communicating class of M_0, let $\mathcal{S}_0$ be all but one representative of that class, and let $M_\epsilon = \begin{pmatrix} \widetilde{M}_\epsilon & \overline{N}_\epsilon \\ \widetilde{N}_\epsilon & (M_\epsilon)_{s,s} \end{pmatrix}$ by partitioning M_ϵ according to $\mathcal{S}_0$.

Theorem 7.29. *If $\left(\widehat{M}_\epsilon, p_\epsilon, i_\epsilon\right)$ is the reduction with respect to $\mathcal{S}_0$ of M_ϵ, then $\widehat{M}_\epsilon \sim \widehat{M}'_\epsilon$, where $\widehat{M}'_\epsilon$ is the unique perturbed Markov matrix whose off-diagonal entries equal those of $M'_\epsilon = \widetilde{M}_\epsilon + \overline{N}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_\epsilon$.*

Proof. For concreteness, let $|\mathcal{S}| = m$, with state $n - m$ as the chosen representative. First, observe that by Lemma 7.22 and Theorem 7.28, we may assume that the entries of $R\left((M_\epsilon)_{s,s}\right)$ are either 0 or ∞. By Lemma 7.26, the entries of $R\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)$ are also either 0 or ∞. In particular, $R\left(\left(I - \overline{M}_0\right)^{-1}\right) = R\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)$ and $R\left(\widehat{M}_\epsilon\right) = R\left(M'_\epsilon\right)$. Therefore, $\mathcal{G}\left(\widehat{M}_\epsilon\right) = \mathcal{G}\left(M'_\epsilon\right)$ and $\mathcal{P}\left(\widehat{M}_\epsilon\right) = \mathcal{P}\left(M'_\epsilon\right)$. Since $\widehat{M}_\epsilon$ is regular, we may apply Lemma 7.21 to obtain $\widehat{M}'_\epsilon$. □

7.7 The SSD Algorithm

In this section, we present our algorithm for computing the SSD of a PMM and prove that it is correct. Intuitively, given a PMM, M_ϵ, the algorithm is as follows:

1. Examine the corresponding unperturbed, Markov matrix, M_0; this corresponds to line 2 in Algorithm 2.

2. If it is unichain, then, as we will shortly observe, its unique stable distribution is precisely the SSD of M_ϵ, so we are done; this corresponds to lines 3-4 in Algorithm 2.

3. Otherwise, take a maximal reduction of M_ϵ, i.e., reduce each of its communicating classes to a singleton; this corresponds to line 5-6 in Algorithm 2.

4. Then apply the non-uniform scaling construction of Lemma 7.16, and iterate (via tail recursion); this corresponds to lines 7-8 in Algorithm 2.

Algorithm 2 To Compute the SSD of a PMM.

```
1: function v_0 = SSD (M_ε) {
2:     C = commClasses (M_0);
3:     if (C.numClosed == 1)
4:         return(stab (M_0));
5:     if (C.nonTrivial > 0)
6:         (M_ε, i, C) = reduce (M_ε, C);
7:     (M_ε, D) = nonUniformScale (M_ε, C);
8:     return(normalize (iD (SSD (M_ε))));
9: }
```

Algorithm 3 To Reduce a PMM.

1: **#define** uniformScale(M) (hasZeroOnDiagonalP (M) : $(I+M)/2$? M)

2: **function** (M_ϵ, i, C) = reduce (M_ϵ, i, C) {

3: $M_\epsilon =$ uniformScale (M_ϵ);

4: $i =$ identityMatrix $(M_\epsilon$.dim());

5: **for** $(c = C$.first(); c.next() $\neq NULL; c = c$.next()) {

6: $s = c$.members().rest();

7: c.setMembers $(c$.members().first());

8: $M_\epsilon = (M_\epsilon)_{\overline{s},\overline{s}} + (M_\epsilon)_{\overline{s},s} \left(I - (M_0)_{s,s}\right)^{-1} (M_\epsilon)_{s,\overline{s}}$;

9: $i = iP_s \begin{pmatrix} I \\ \left(I - (M_0)_{s,s}\right)^{-1} (M_0)_{s,\overline{s}} \end{pmatrix}$;

10: }

11: **return**(M_ϵ, i, C);

12: }

We begin by observing that, by Theorem 7.10 and the comments following, as well as Theorem 7.28, at any step in the algorithm, we may replace M_ϵ by an equivalent PMM without affecting the final result (i.e., the SSD). In particular, we may represent M_ϵ by its pair of resistance and cost matrices. Theorem 7.6 then tells us how to carry out any subsequent algebraic operations (i.e., addition and multiplication of PMMs).

We now prove that our termination condition in step 2 is correct.

Theorem 7.30. *Given* $M_\epsilon \in \mathrm{PMM}(n)$*, if* M_0 *is unichain, then* $\operatorname{stab} M_0 = \{\operatorname{ssd} M_\epsilon\}$.

Proof. Letting $v_\epsilon \equiv \operatorname{stab}(M_\epsilon)$, by definition, $M_\epsilon v_\epsilon = v_\epsilon$ for $\epsilon \succ 0$. Taking limits, we have $M_0 v_0 = v_0$, that is, $\operatorname{ssd} M_\epsilon = v_0 \in \operatorname{stab} M_0$. Since M_0 is unichain, $\operatorname{stab} M_0$ is a singleton, so that $\operatorname{stab} M_0 = \{\operatorname{ssd} M_\epsilon\}$. □

Next, observe how we implicitly appeal to Corollary 7.12 and Theorem 7.29 in step 3, and the corresponding call to **reduce**. In this way, we reduce each communicating class, while only inverting a *constant* matrix. Notice also how in line 9 of Algorithm 3, we accumulate the *unperturbed* inclusion of the reduction, $\imath_0$, in the (real-valued) matrix, i. This is correct by Theorem 7.15. Finally, note that as we iterate through each communicating class of M_ϵ, **reduce** eliminates all the elements of each class *except* the first, updating its set of member to be *only* the first. Thus, it returns an updated communicating class decomposition, corresponding to the reduced result. This means that we do not need to recompute from scratch before the call to **nonUniformScale**.

Finally, Lemma 7.16 guarantees that step 4 is correct. Moreover, the algorithm is guaranteed to make progress to termination. Specifically, Corollary 7.17 guarantees that eventually either the condition of line 3 or line 5 holds. Thus, either we terminate immediately or we reduce the dimension of M_ϵ, guaranteeing that we will terminate eventually.

Part III

Two Related Algorithms

Chapter 8

GraphRank

In this chapter, we discuss the problem of ranking and frame a solution in terms of a multi-objective linear program. We then show how to compute Pareto optimal solutions, as well as suggest how to find socially optimal solutions. We will assume we are given a set of individuals and for selected ordered pairs, (i, j), a positive value, $d_{i,j} \geq 0$, indicating that j is superior to i by $d_{i,j}$ units. For example, if the individuals are sports teams, $d_{i,j}$ may represent the positive differential in points scored or games won (e.g., team j won $d_{i,j}$ more times against team i than it lost). We can represent this data naturally as a weighted, directed graph, $G = (V, E, d, s, e)$, where the vertices, V, correspond to the individuals being ranked, the edges, E, correspond to the set of comparisons, and $s(\alpha), e(\alpha) : E \to V$ are the starting and ending vertices of α, respectively. We will assume that V and E are finite. The objective is then to determine a ranking function, $x : V \to \mathbb{R}$ which is consistent with the given data, as well as optimal, in some sense.

If the corresponding graph is not connected, then vertices in different components are necessarily incomparable, so we will assume that G is connected. If the corresponding *undirected* graph is connected and acyclic, then there is a unique solution to the corresponding system of equations, $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = d_\alpha,\ \forall\, \alpha \in E$, up to a constant shift. In general, this ranking problem is complicated by the existence of undirected cycles in the graph which lead to an inconsistent system of equations. While there are general techniques for solving such over-determined systems of linear equations, we would prefer a technique which is motivated by this specific application.

One approach taken in the literature for undirected graphs is to simply throw out the

smallest number of edges to obtain a directed acyclic graph (DAG) and apply standard topological sort techniques to the result (Kenyon-Mathieu and Schudy, 2007; Ailon, 2007). This has the advantage of avoiding ties, but does not address how to *optimally* sort individuals in the resulting graph. Thus, we take an alternative approach which will lead to a non-trivial ranking whenever possible. We give one justification for this approach here and address it in further detail in Section 8.3.

For a ranking to be consistent with the data, we must at least have $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha,\ \forall\, \alpha \in E$. In other words, if the data indicates that team i is better than team j, it should be ranked no worse. We wish to create an "informative" ranking, by distinguishing individuals from one another whenever possible. It is easy to show that these constraints imply that any feasible ranking must be constant on directed cycles. That is, with these constraints it is impossible to use a feasible ranking to distinguish between individuals within a directed cycle. However, by collapsing all cycles to single vertices, we obtain a DAG and we can sort the vertices in strictly increasing order. Therefore, topological sort on unweighted graphs can be viewed as an optimization problem, where we are trying to minimize the number of strict equalities in our constraints. This objective has the nice property of treating the data "fairly" by handling all the given edges in the same manner.

However, a ranking should not overstate the degree of superiority of team i over j. That is, we should impose the additional constraints $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha,\ \forall\, \alpha \in E$. Now instead of distinguishing individuals *whenever* possible, we can also seek to do so as *much* as possible. That is, we want a ranking which is maximal with respect to the multiple objectives, $f_{s,t}(x) = x(t) - x(s)$, such that $(s,t) = (s(\alpha), e(\alpha))$ with $\alpha \in E$. Notice that if we consider unweighted graphs as the case when $d_\alpha \equiv 1$, this will automatically minimize the number of tight left-hand side constraints, so this is a natural generalization of the previous problem.

Formally, given a weighted, directed graph, $G = (V, E, d)$, we restrict attention to the set of valuations on G, $C(G)$, satisfying the corresponding system of linear constraints:

$$C(G) = \left\{x : V \to \mathbb{R} \mid 0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha,\ \forall\, \alpha \in E\right\}.$$

Observe that $C(G) \neq \emptyset$, since it contains every constant function on V. Now define a pre-order (with corresponding partial order and equivalence relation) on $C(G)$, $\preceq$, so that $x \preceq x'$ iff $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq x'\left(e(\alpha)\right) - x'\left(s(\alpha)\right), \forall\, \alpha \in E$. Our goal is then to

solve the multi-objective, linear program, **Rank**(G): compute $x^* \in \max_{C(G)}$, that is, find a Pareto optimal, feasible ranking.

This type of system of constraints has been well-studied in operations research, particularly in the theory of scheduling (Corman et al., 2001). The graph is then called a PERT (Program Evaluation and Review Technique) chart. However, the classical problem generally involves *minimizing* the difference between the largest and smallest values of x, which in our case would yield the trivial ranking, $x(v) \equiv 0$. However, we might wish to refine our search to a Pareto optimal solution which is socially optimal with respect to some aggregation function, such as $\max_{s,\, t \in V} x(t) - x(s)$ or $\sum_{\alpha \in E} x(e(\alpha)) - x(s(\alpha))$.

In the latter case, the problem may be recast as a standard linear program, where a solution may always be taken at a vertex of $G(G)$, corresponding to a consistent collection of tight constraints. Intuitively, the tight constraints of the form $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = d_\alpha$ correspond to a choice of edges, while the remaining equality constraints, $x\left(e(\alpha)\right) = x\left(s(\alpha)\right)$ should only be those forced by the geometry of G. This suggests that we pursue a general approach to finding Pareto optimal solutions corresponding to maximizing the difference on selected edges, using the geometry of G to infer the subsequent equality constraints.

We will carry out this program in Section 8.2 and give an algorithm for computing a Pareto optimal solution. We then discuss related work and future directions for research. We conclude by outlining applications of these techniques to voting and information retrieval.

8.1 Existence of Solutions

In this section, we discuss some background results related to the problem, **Rank**(G), introduced in the previous section. Most importantly, we will show that solutions exist, and that feasible rankings are constant on strongly connected components (SCCs) of G. SCCs are most easily described in terms of the natural preorder of the vertices, given by the "leads to" relation, $\rightsquigarrow$, where $i \rightsquigarrow j$ iff there is a directed path in G from i to j. Since G is connected, for each $s, t \in V$, there is a sequence $\{v_i\}_{i=0}^k \subset V$ such that $s \rightsquigarrow v_0 \reflectbox{$\rightsquigarrow$} v_1 \rightsquigarrow \cdots \reflectbox{$\rightsquigarrow$} v_k \rightsquigarrow t$. This relation defines an equivalence relation, $\sim$, where $i \sim j$ iff $i \rightsquigarrow j$ and $j \rightsquigarrow i$, a corresponding partial order on the set of equivalence classes, which we will

denote by $\preceq$, and a *strict* partial order, $\prec$. The strongly connected components of G are simply equivalence classes with respect to $\sim$. They are also often called "communicating" classes.

Given an *undirected* walk in G, we may define its length by summing the weights of the edges, where we weight a forward edge as d_α and a reversed edge as 0. Let $l_G(s,t)$ be the length of the shortest such walk from s to t. Since G is connected, this set of walks is non-empty, so that $l_G(s,t)$ is well-defined for all $s,t \in V$.

Lemma 8.1. *If $x \in C(G)$, $x(t) - x(s) \leq l_G(s,t)$. In particular, if $s \rightsquigarrow t$, $x(s) \leq x(t)$.*

Proof. Consider any undirected walk from s to t. For each forward edge, we have $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha$. For each reversed edge, we have $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right)$, so that $x\left(s(\alpha)\right) - x\left(e(\alpha)\right) \leq 0$. When we sum these inequalities, the left-hand side telescopes to $x(t) - x(s)$, while the right-hand side yields the length of the undirected walk. Since this inequality holds for all such walks, it holds for the shortest such one, i.e., when the right-hand side is $l_G(s,t)$. If $s \rightsquigarrow t$, then there is an undirected walk from t to s consisting entirely of reversed edges, so that $l_G(t,s) = 0$, $x(s) - x(t) \leq 0$, and $x(s) \leq x(t)$. □

We can now prove that **Rank**(G) has at least one solution. For the following, we will assume only that there is a well-defined operation, I, taking a directed graph to pairs of its vertices, $I(G) \subset V \times V$, and that $x \preceq x'$ iff $x(t) - x(s) \leq x'(t) - x'(s), \forall\, (s,\, t) \in I(G)$. For example,

$$I(G) = \{(s,\, t) \mid \deg_{in}(s) = \deg_{out}(t) = 0\} \quad \text{or} \quad \{(s,\, t) \mid s = s(\alpha),\, t = e(\alpha),\, \alpha \in E\}\,.$$

Lemma 8.2. *If $x' - x$ is constant, then $x \sim x'$. Moreover, if $x \in C(G)$, then $x' \in C(G)$.*

Proof. If $x' = x + c$, $x'\left(t'\right) - x'\left(s'\right) = x\left(t'\right) - x\left(s'\right),\ \forall\, s', t' \in V$. Since $\sim$ and $C(G)$ are defined solely in terms collections of such differences, the conclusions of the Lemma follow. □

Theorem 8.3. $\max C(G) \neq \emptyset$.

Proof. Let $M = \max_{s,t \in V} l_G(s,t)$. By Lemma 8.1 $-l_G(v,u) \leq x(u) - x(v) \leq l_G(u,v)$, so that $|x(u) - x(v)| \leq \max\left\{l_G(u,v), l_G(v,u)\right\} \leq M$. By Lemma 8.2, there is a representative of each equivalence class of $C(G)$ in

$$C'(G) \equiv \left\{ x \in C(G) \mid \sum_{v \in V} x(v) = 0 \right\}.$$

For any $v \in V$, if $N = |V|$, $|x(v)| = |x(v)| = \frac{1}{N}\left|\sum_{u \in V} x(u) - x(v)\right| \leq \frac{1}{N}\sum_{u \in V}|x(u) - x(v)| \leq \frac{1}{N}NM = M$. Thus, $C'(G)$ is closed and bounded, hence compact. Consider

$$D(x) \equiv \sum_{(s,t) \in I(G)} x(t) - x(s).$$

This is continuous on $C'(G)$ and hence attains a maximum there at some x^*. This must be maximal in $C(G)$, for if $x \prec x'$, we can assume without loss of generality that $x' \in C'(G)$, so that $D(x) < D\left(x'\right)$, which is a contradiction. □

Observe that when G is strongly connected, the constant solution is the unique (up to constant shift) solution to **Rank**(G). More generally, any feasible ranking is constant on strongly connected components of G.

Theorem 8.4. *If $x \in C(G)$ and s and t are in a common cycle (i.e., $s \sim t$), then $x(s) = x(t)$. In particular, $x(v)$ is constant on strongly connected components.*

Proof. Since $s \sim t$, $s \preceq t$, and $x(s) \leq x(t)$, by Lemma 8.1. Likewise, $x(t) \leq x(s)$, so that $x(s) = x(t)$ for s and t within the same strongly connected component. □

8.2 Computing Pareto Optimal Solutions

In this section we show how to solve **Rank**(G) from the Introduction by using four basic operations, reversing 0 weight edges, collapsing cycles, dropping multiple edges, and "shifting" vertices, to recursively simplify the problem. In so doing, we give a constructive proof of existence of solutions to **Rank**(G). We begin by showing how we may reduce the problem on an arbitrary graph to a corresponding problem on a simple, directed acyclic graph without 0 edges, obtained by reversing 0 edges, collapsing cycles and then dropping all but the smallest of multiple edges.

Given a a weighted, directed graph, $G = (V, E, d, s, e)$, let $E_0 = \{\alpha \in E \mid d_\alpha = 0\}$ be the set of 0 weight edges. We may define a new graph $T(G) = (V', E', d', s', e')$ with $V' \equiv V$, where E' is essentially the same as E, except we introduce the reverse of the edges in E_0. Specifically, $E' \equiv E \times \{0\} \cup E_0 \times \{1\}$, with $d'_{(\alpha,0)} \equiv d_\alpha$, $d'_{(\alpha,1)} \equiv 0$, $s'(\alpha, 0) \equiv s(\alpha)$, $s'(\alpha, 1) \equiv e(\alpha)$, $e'(\alpha, 0) \equiv e(\alpha)$, and $e'(\alpha, 1) \equiv s(\alpha)$. Notice that $C(G) = C(T(G))$, since we are simply adding redundant equality constraints.

Theorem 8.5. *The set of feasible rankings for G and $T(G)$ are identical, that is, $C(G) = C(T(G))$. Likewise, $x \preceq_G x'$ iff $x \preceq_{T(G)} x'$. In particular,* $\max C(T(G)) = \max C(G)$.

Proof. Given $x \in C(G)$ and $\alpha' = (\alpha, i) \in E'$, there are two cases to consider. If $i = 0$, then $\alpha \in E$ and $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha$. Since $s'(\alpha, 0) = s(\alpha)$, $e'(\alpha, 0) = e(\alpha)$, and $d'_{(\alpha,0)} = d_\alpha$, $0 \leq x\left(e'(\alpha, 0)\right) - x\left(s'(\alpha, 0)\right) \leq d_{(\alpha,0)}$. If $i = 1$, then $\alpha \in E_0$ and $d_\alpha = 0$. Therefore, $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha = 0$ and $0 \leq x\left(s(\alpha)\right) - x\left(e(\alpha)\right) \leq 0 = d_{(\alpha,1)}$. Since $s'(\alpha, 1) = e(\alpha)$ and $e'(\alpha, 1) = s(\alpha)$, $0 \leq x\left(e'(\alpha, 1)\right) - x\left(s'(\alpha, 1)\right) \leq d_{(\alpha,1)}$. Therefore, $x \in C(T(G))$.

Conversely, if $x \in T(G)$ and $\alpha \in E$, then $0 \leq x\left(e'(\alpha, 0)\right) - x\left(s'(\alpha, 0)\right) \leq d_{(\alpha,0)}$. As before, $s'(\alpha, 0) = s(\alpha)$, $e'(\alpha, 0) = e(\alpha)$, and $d'_{(\alpha,0)} = d_\alpha$, so that $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha$ and $x \in C(G)$.

Now assume that $x_i \in C(G) = C(T(G))$, $i = 1, 2$, and $x_1 \preceq_G x_2$. Given $\alpha' = (\alpha, i) \in E'$, there are again two cases to consider. If $i = 0$, then $\alpha \in E$, and since $x_1 \preceq_G x_2$, $x_1\left(e(\alpha)\right) - x_1\left(s(\alpha)\right) \leq x_2\left(e(\alpha)\right) - x_2\left(s(\alpha)\right)$, so that $x_1\left(e'(\alpha, 0)\right) - x_1\left(s'(\alpha, 0)\right) \leq x_2\left(e'(\alpha, 0)\right) - x_2\left(s'(\alpha, 0)\right)$. If $i = 1$, then $\alpha \in E_0$, $x_i\left(e(\alpha)\right) = x_i\left(s(\alpha)\right)$, and $x_1\left(e'(\alpha, 1)\right) - x_1\left(s'(\alpha, 1)\right) = 0 \leq 0 = x_2\left(e'(\alpha, 1)\right) - x_2\left(s'(\alpha, 1)\right)$. Thus, $x_1 \preceq_{T(G)} x_2$.

Conversely, assume that $x_1 \preceq_{T(G)} x_2$. Given $\alpha \in E$, then $\alpha' = (\alpha, 0) \in E'$. Since $x_1 \preceq_{T(G)} x_2$, $x_1\left(e'(\alpha, 0)\right) - x_1\left(s'(\alpha, 0)\right) \leq x_2\left(e'(\alpha, 0)\right) - x_2\left(s'(\alpha, 0)\right)$. Therefore, $x_1\left(e(\alpha)\right) - x_1\left(s(\alpha)\right) \leq x_2\left(e(\alpha)\right) - x_2\left(s(\alpha)\right)$, so that $x_1 \preceq_G x_2$. □

Given a a weighted, directed graph, $G = (V, E, d, s, e)$, let $[v] = \left\{v' \in V \mid v' \sim v\right\}$. We wish to define a graph $P(G) = (V', E', d', s', e')$, where $V' = \{[v] \mid v \in V\}$. That is, the vertices of $P(G)$ correspond to the strongly-connected components of G. Since strongly-connected components are unions of cycles, we have effectively collapsed each cycle to a single vertex. To guarantee that the result is simple, we discard all loops and retain only the smallest weight edge between any two vertices. Specifically, given $s', t' \in V'$ with $s' = [s]$ and $t' = [t]$, such that $s \prec_G t$ and $E_{s',t'} \equiv \left\{\alpha \in E \mid s\left(\alpha\right) \in s', e\left(\alpha\right) \in t'\right\} \neq \emptyset$, choose $\alpha_{s',t'} = \arg\min_{\alpha \in E_{s',t'}} d_\alpha$, and take E' to be the set of all such edges. Finally, take $d' = d|_{E'}$, $s'(\alpha) = [s(\alpha)]$, and $e'(\alpha) = [e(\alpha)]$. Notice that by construction, $s'\left(\alpha_{s',t'}\right) = s'$ and $e'\left(\alpha_{s',t'}\right) = t'$.

Now observe that there is a natural, order-preserving correspondence between $C(G)$ and $C(P(G))$.

Theorem 8.6. *The mapping p given by $p(x)(v) = x([v])$ is a bijective, order preserving correspondence between $C(P(G))$ and $C(G)$. In particular, $p(\max C(P(G))) = \max C(G)$.*

Proof. We first show that, if $x' \in C(P(G))$, $x \equiv p\left(x'\right) \in C(G)$. Given $\alpha \in E$, we must show that $0 \leq x\left(e(\alpha)\right) - x\left(s(\alpha)\right) \leq d_\alpha$. By definition, $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = x'([e(\alpha)]) - x'([s(\alpha)])$. If $s(\alpha) \sim e(\alpha)$, this is 0 and we are done. Otherwise, letting $t' \equiv [e(\alpha)])$ and $s' \equiv [s(\alpha)]$, we have $s(\alpha) \prec_G e(\alpha)$ and $\alpha \in E_{s',t'}$, so that $x'([e(\alpha)]) - x'([s(\alpha)]) = x'\left(t'\right) - x'\left(s'\right) = x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right)$. Since $x' \in C(P(G))$, $0 \leq x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right) \leq d_{\alpha_{s',t'}} \leq d_\alpha$. Since $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right)$, we have shown that $x \in C(G)$.

Now observe that, by Theorem 8.4, any $x \in C(G)$ is constant on equivalence classes, $[v]$, so that the inverse mapping, $p^{-1}\left(x\right)\left([v]\right) = x(v)$, is well-defined. If $x' = p^{-1}\left(x\right)$, we must show that $x' \in C(P(G))$. If $\alpha \in E'$, then $\alpha = \alpha_{s',t'}$ with $s'(\alpha) = s'$ and $e'(\alpha) = t'$. Moreover, $s' = [s(\alpha)]$ and $t' = [e(\alpha)]$ and $x'\left(e'\left(\alpha\right)\right) - x'\left(s'\left(\alpha\right)\right) = x'\left([e(\alpha)]\right) - x'\left([s(\alpha)]\right) = x\left(e(\alpha)\right) - x\left(s(\alpha]\right)$, which is between 0 and $d_\alpha = d'_\alpha$, since $x \in C(G)$.

Finally, it remains to show that p and p^{-1} are order-preserving functions. Assume that $x_1 \preceq x_2$ with $x_i \in C(P(G))$, and consider $\alpha \in E$. For convenience, let $t = e(\alpha)$, $s = s(\alpha)$, and observe that $p\left(x_i\right)\left(t\right) - p\left(x_i\right)\left(s\right) = x_i([t]) - x_i([s])$. There are two cases to consider. Either $s \sim_G t$ or $s \prec_G t$. In the first case, $x_i([t]) = x_i([s])$, we have $p\left(x_1\right)\left(t\right) - p\left(x_1\right)\left(s\right) = 0 \leq 0 = p\left(x_1\right)\left(t\right) - p\left(x_1\right)\left(s\right)$. Otherwise, $s \prec_G t$ and $\alpha \in E_{[s],[t]}$, so that $\alpha\_[s], [t] \in E'$ with $[s] = s\left(\alpha_{[s],[t]}\right)$ and $[t] = e\left(\alpha_{[s],[t]}\right)$. Since $x_1 \preceq x_2$, $x_1([t]) - x_1([s]) \leq x_2([t]) - x_2([s])$, or equivalently, $p\left(x_1\right)\left(t\right) - p\left(x_1\right)\left(s\right) \leq p\left(x_1\right)\left(t\right) - p\left(x_1\right)\left(s\right)$. Therefore, $p\left(x_1\right) \leq p\left(x_2\right)$.

Conversely, assume that $x'_1 \preceq x'_2$ with $x'_i \in C(G)$. Let $x'_i = p\left(x_i\right)$ and consider $\alpha \in E'$. Then $\alpha = \alpha_{[s],[t]}$, with $s(\alpha) = s$ and $e(\alpha) = t$. Since $s'(\alpha) = [s]$ and $e'(\alpha) = [t]$, we must show that $x_1([t]) - x_1([s]) \leq x_2([t]) - x_2([s])$. Since $x'_1 \preceq x'_2$, $x'_1(t) - x'_1(s) \leq x'_2(t) - x'_2(s)$. By definition, $x'_i(t) = x_i([t])$ and $x'_i(s) = x_i([s])$. Therefore, $x_1([t]) - x_1([s]) \leq x_2([t]) - x_2([s])$, as desired, so that $x_1 \preceq x_2$. □

We now introduce a novel operation to "shift" a vertex. This corresponds to feasible pivot in a simplex tableau for $C(G)$, or equivalently, a change of variables that forces $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = d_\alpha$. Specifically, given a weighted, directed graph without self-loops,

$G = (V, E, d, s, e)$, we will say that an edge, $\alpha \in E$, is *feasible* iff $d_\alpha = \min_{e(\beta)=e(\alpha)} d_\beta$. For any feasible edge, α, we then define a graph $R_\alpha(G)$ which is identical to G, but with a new set of edge weights. Specifically, let $R_\alpha(G) = (V, E, d', s, e)$ with

$$d'_\beta \equiv \begin{cases} d_\beta - d_\alpha, & \text{if } e(\beta) = e(\alpha) \\ d_\beta + d_\alpha, & \text{if } s(\beta) = e(\alpha) \\ d_\beta, & \text{otherwise} \end{cases} .$$

As before, there is a corresponding mapping, r_α, taking valuations on $R_\alpha(G)$ to valuations on G.

Theorem 8.7. *Consider the mapping r_α, given by*

$$r_\alpha\left(x'\right)(v) \equiv \begin{cases} x'(v) + d_\alpha, & \textit{if } v = e(\alpha) \\ x'(v), & \textit{otherwise} \end{cases}$$

This is bijective, order-preserving correspondence between:

$$\operatorname{domain} r_\alpha \equiv \left\{x' \in C(R(G)) \mid x'(e(\beta)) - x'(s(\beta)) \geq d_\alpha, \textit{ if } e(\alpha) = s(\beta)\right\}$$

and

$$\operatorname{im} r_\alpha = \left\{x \in C(G) \mid x(e(\beta)) - x(s(\beta)) \geq d_\alpha, \textit{ if } e(\alpha) = e(\beta)\right\}.$$

Proof. We first show that $x \equiv r_\alpha\left(x'\right) \in C' \equiv \{x \in C(G) \mid x(e(\beta)) - x(s(\beta)) \geq d_\alpha, \text{ if } e(\alpha) = e(\beta)\}$, for any $x' \in \operatorname{domain} r_\alpha$. Given $\beta \in E$, by assumption, $0 \leq x'\left(e\left(\beta\right)\right) - x'\left(s\left(\beta\right)\right) \leq d'_\beta$, with $d_\alpha \leq x'\left(e\left(\beta\right)\right) - x'\left(s\left(\beta\right)\right)$, if $e(\alpha) = s(\beta)$. If $e(\beta) = e(\alpha)$, then $d'_\beta = d_\beta - d_\alpha$, $x'\left(e(\beta)\right) = x\left(e(\beta)\right) - d_\alpha$, and $x'\left(s(\beta)\right) = x\left(s(\beta)\right)$, since $s(\beta) \neq e(\beta) = e(\alpha)$. Therefore, $0 \leq d_\alpha \leq x\left(e(\beta)\right) - x\left(s(\beta)\right) \leq d_\beta$. In particular, $x\left(e(\alpha)\right) - x\left(s(\alpha)\right) = d_\alpha$.

Similarly, if $s(\beta) = e(\alpha)$, then $d'_\beta = d_\beta + d_\alpha$, $x'\left(s(\beta)\right) = x\left(s(\beta)\right) - d_\alpha$, and $x'\left(e(\beta)\right) = x\left(e(\beta)\right)$, since $e(\beta) \neq s(\beta) = e(\alpha)$. Since $d_\alpha \leq x'\left(e\left(\beta\right)\right) - x'\left(s\left(\beta\right)\right) \leq d'_\beta$, $0 \leq x\left(e(\beta)\right) - x\left(s(\beta)\right) \leq d_\beta$.

Finally, if $e(\alpha) \neq s(\beta), e(\beta)$, then $d'_\beta = d_\beta$, $x'\left(s(\beta)\right) = x\left(s(\beta)\right)$, and $x'\left(e(\beta)\right) = x\left(e(\beta)\right)$. Thus, $0 \leq x\left(e(\beta)\right) - x\left(s(\beta)\right) \leq d_\beta$ for all $\beta \in E$, with $d_\alpha \leq x(e(\beta)) - x(s(\beta))$, when $e(\alpha) = e(\beta)$. In other words, $x \in C'$, so that $\operatorname{im} r_\alpha \subset C'$. Conversely, if $x \in C'$, we may then define $x'(v) = x(v) - d_\alpha$, if $v = e(\alpha)$, and $x(v)$, otherwise. The previous calculations may all clearly be reversed, to show that $x' \in \operatorname{domain} r_\alpha$.

We now observe that both r_α and r_α^{-1} are order-preserving. If $x \equiv r_\alpha\left(x'\right)$, then for any $s, t \in V$, notice that $x_i(t) - x_i(s) = r_\alpha\left(x_i\right)(t) - r_\alpha\left(x_i\right)(s) - c$, where $c = -d_\alpha$, d_α, or 0, depending on whether s and/or t equal $e(\alpha)$. In any case, $x_1(t) - x_1(s) \leq x_2(t) - x_2(s)$ iff $r_\alpha\left(x_1\right)(t) - r_\alpha\left(x_1\right)(s) \leq r_\alpha\left(x_2\right)(t) - r_\alpha\left(x_2\right)(s)$, for all $s, t \in V$. In particular, $x_1 \preceq x_2$ iff $r_\alpha\left(x_1\right) \preceq r_\alpha\left(x_2\right)$. □

Lemma 8.8. *If $x' \preceq x \in C(G)$ for $x' \in \operatorname{im} r_\alpha$, then $x \in \operatorname{im} r_\alpha$.*

Proof. Assume that $x' \preceq x \in C(G)$, $x' \in \operatorname{im} r_\alpha$. Then, for any $\beta \in E$, $x(e(\beta)) - x(s(\beta)) \geq x'(e(\beta)) - x'(s(\beta))$. By Theorem 8.7, if $e(\alpha) = e(\beta)$, $x'(e(\beta)) - x'(s(\beta)) \geq d_\alpha$. Therefore, $x(e(\beta)) - x(s(\beta)) \geq d_\alpha$, and again by Theorem 8.7, $x \in \operatorname{im} r_\alpha$. □

Lemma 8.9. $r_\alpha\left(\max C\left(R_\alpha(G)\right) \cap \operatorname{domain} r_\alpha\right) \subset \max C(G)$.

Proof. Let $x_1 \in \max C\left(R_\alpha(G)\right) \cap \operatorname{domain} r_\alpha$ and $r_\alpha\left(x_1\right) \preceq x \in C(G)$. By Lemma 8.8, $x = r_\alpha\left(x_2\right)$, for $x_2 \in \operatorname{domain} r_\alpha$. Therefore, by Theorem 8.6, $x_1 \preceq x_2$, which implies that $x_1 \sim x_2$, since a_1 is maximal. Thus, $r_\alpha\left(x_1\right) \sim r_\alpha\left(x_2\right) = x$, again by Theorem 8.6. In particular, $r_\alpha\left(x_1\right) \in \max C(G)$. □

Combining these results gives Algorithm 4 which computes Pareto optimal rankings.

Algorithm 4 To Solve **Rank**(G).

```
1: function x = rk (G) {
2:     G' = P(T(G))
3:     if (G'.numVertices == 1)
4:         return(p(0));
5:     Choose a feasible α ∈ E such that e(α) is maximal.
6:     G' = R_α(G')
7:     return(r_α(p(rk(G'))));
8: }
```

Theorem 8.10. *Algorithm 4 produces a solution to **Rank**(G).*

Proof. By construction, $R_\alpha\left(G'\right)$ has a 0 weight edge, so that $T\left(R_\alpha\left(G'\right)\right)$ will have at least one fewer strongly connected components than G. Thus, after the initial call, line 2 decreases the number vertices by at least 1. In particular, we are eventually left with a trivial graph, containing a single vertex and no edges. Thus, the algorithm is guaranteed to terminate.

If G' is trivial, by Lemma 8.2, the constant ranking $x(v) = 0$ is clearly the unique feasible ranking, up to equivalence. Moreover, Theorems 8.6 and 8.5 imply that p takes this optimal ranking for G' to an optimal ranking for G. Moreover, by line 5, G' is a simple, acyclic directed graph. Thus, there can be no edges, β, such that $e(\alpha) = s(\beta)$. In particular, $\operatorname{domain} r_\alpha = C(R(G))$, so that we may apply Lemma 8.9, along with Theorems 8.6 and 8.5 to prove that line 7 is correct. □

Notice that after the initial call to $\mathrm{rk}(G)$, we may optimize the subsequent calls to P and T, since we know precisely those edges which became 0 in G, and there will only be one non-trivial strongly connected component of $T(G)$. In practice, we might expect these calls to be of roughly constant cost, so that the entire algorithm should be linear in the size of G.

8.3 Related Work

The problem of ranking from a directed graph has been studied. However, the approach has been to simply throw out the smallest cumulative weight of edges to obtain a directed acyclic graph (DAG), and apply standard topological sort techniques to the result (Kenyon-Mathieu and Schudy, 2007; Ailon, 2007). While this approach has generated a lot of interest, since it is an NP-hard problem to solve exactly, one could argue, however, that throwing away edges is not realistic. For example, if G consists of a 3-cycle with equally weighted edges, while throwing away an edge will lead to an ordering of the vertices, the resulting order will depend crucially on which edge is thrown away, even though the edge can only be chosen "randomly". The value of such a strict ordering is dubious, since the available evidence indicates that the vertices are indistinguishable.

When the edges are weighted, this approach becomes even more problematic. Consider the case where $d_{a,b} = 1$, $d_{b,c} = 2$, and $d_{c,a} = 3$. If we throw out smallest edge, (a, b), we obtain the ordering $b < c < a$, even though the data indicates that a and b are the most similar. If we throw out largest edge, (c, a), the situation is even worse, since we obtain the ordering $a < b < c$, although the data indicates that a is superior to c by the widest margin.

Moreover, this approach only reduces the problem to ranking a DAG, and does not address how to *optimally* sort individuals in the resulting graph. Thus, we studied an alternative which leads to a non-trivial ranking whenever possible, that is, when the graph is

not strongly connected with non-zero weight edges between the strongly connected components. In practice, if a given statistic leads to a strongly connected graph, we suggest that we should look for different statistics to reduce the number of cycles until the result is not strongly connected. Consider our original example of sports teams. An initial approach might be to simply weight an edge from i to j by number of games won by j over i. But in sports where teams play each other many times, this will produce a 2-cycle between almost every team. By combining the number of wins and the number of losses as the difference, we obtain a graph with many fewer cycles. If necessary, we can employ additional statistics to break "ties" in the resulting ranking.

8.4 Applications

Another interesting example of statistical aggregation comes from the field of information retrieval. Ailon (2007) considers the problem of aggregating partial rankings produced by different Web search engines. Each partial ranking gives a graph of pairwise comparisons between consecutive members of each ranking. We may then average the graphs together, weighted according to our belief in the quality of the results of each engine, to obtain a directed graph for which **Rank**(G) may give useful results.

Conjecture: Alternatively, if we are given cardinal ranking and average the corresponding tournament graphs, we obtain a graph where the weights reflect the number of head-to-head wins, on which we may then solve **Rank**(G). This ranking scheme will then satisfy the Condorcet criteria.

8.5 Conclusion

In this paper, we discuss the problem of ranking and frame a solution in terms of a multi-objective linear program. We then show how to compute Pareto optimal solutions, as well as suggest how to find socially optimal solutions. We then suggest a modified version of the algorithm which we believe will compute corresponding socially optimal rankings.

Chapter 9

QuickRank

9.1 Introduction

A fundamental problem in the field of social network analysis is to rank individuals in a society according to their implicit "importance" (e.g., power or influence), derived from a network's underlying topology. More precisely, given a social network, the goal is to produce a (cardinal) *ranking*, whereby each individual is assigned a nonnegative real value, from which an ordinal ranking (an ordering of the individuals) can be extracted if desired. In this paper, we propose a solution to this problem specifically geared toward social networks that possess an accompanying hierarchical structure.

A social network is typically encoded in a *link graph*, with individuals represented by vertices and relationships represented by directed edges, or "links," annotated with weights. Given a link graph, there are multiple ways to assign meaning to the weights. On one hand, one can view the weight on a link from i to j as expressing the distance from i to j—a quantity inversely related to j's importance. On the other hand, one can view each weight as the level of endorsement, or respect, i grants j—a quantity directly proportional to j's importance. We adopt this latter interpretation.

Under either interpretation (weights as distances or weights as endorsements), a social network can be seen as a collection of judgments, one made by each individual in the society. Correspondingly, we seek a means of aggregating individual judgments into a single collective ranking. In other words, we consider the aforementioned fundamental problem in social network analysis as akin to a key question in voting: how to aggregate

the preferences of many individuals into a single collective persuasion that reflects the preferences of the population as a whole.

Given a link graph, perhaps the most basic ranking scheme is degree centrality, in which i's rank is a combined measure of its indegree, the strength of the endorsements i receives, and outdegree, the strength of the endorsements i makes. It is straightforward to compute this metric. However, it could be argued that it is also sensible to take into account inferred endorsements: e.g., if i endorses j and j endorses k, then i endorses k in a sense. At the opposite end of the spectrum lie ranking schemes that incorporate all such inferred endorsements.

Central to these alternatives is a hypothesis due to Bonacich (1972): *an individual is deemed important if he is endorsed by other important individuals*. In other words, the strength of an endorsement should be construed relative to the rank of the individual making the endorsement. In terms of our voting analogy, Bonacich suggests relating the collective ranking to the sum of all individual judgments, each weighted by its respective rank as determined by the collective. The fixed point of this averaging process—the principal eigenvector of the link graph—defines Bonacich's metric, also known as eigenvector centrality. Although intuitively appealing, the computation of this fixed point can be prohibitive in large networks.

Recently, computer scientists have developed related schemes to rank web pages based on the Web's underlying topology. Viewed as a social network, web pages are individuals and hyperlinks are links. The most prominent approach to ranking web pages is the PageRank algorithm (Page and Brin, 1998; Page et al., 1998), upon which the Google search engine is built. PageRank aggregates the information contained in the Web's hyperlinks to generate a ranking using a process much like Bonacich's method for computing eigenvector centrality.

In this paper, we present QuickRank, an efficient algorithm for computing a ranking in an *hierarchical social network*. Many social networks are hierarchical. One apt example already mentioned is the Web, where the individuals are web pages, the network structure is provided by hyperlinks from one web page to another, and an explicit hierarchical structure is given by the Web's domains, subdomains, and so on. Another fitting example is the Enron email database, where individuals are employees, the network structure is given by emails from one employee to another, and an explicit hierarchical structure is given by the

corporate hierarchy. Yet another compelling example is a citation index. In this case, the individuals are publications, the network structure is dictated by the references from one publication to another, and an explicit hierarchical structure is given by the categorization of publications by fields (e.g., computer science), subfields (e.g., AI, theory, and systems), and so on.

As we sketch the key ideas behind the QuickRank algorithm in this introductory section, we allude to the sample hierarchical social network shown in Figure 9.1, a network of web pages within a domain hierarchy. The web pages, indicated by gray rectangles, are the individuals in this society. Social relationships between these individuals (i.e., hyperlinks between web pages) are shown as dashed lines with arrows. The domain hierarchy is drawn using solid lines with domains and subdomains as interior nodes, indicated by solid black circles, and web pages as leaves (gray rectangles).

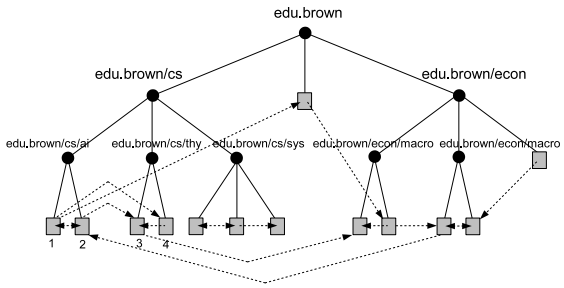


Figure 9.1: A sample hierarchical social network.

Up to normalization, a ranking is a probability distribution. Given any normalized ranking (i.e., probability distribution) of the individuals in an hierarchical social network, by conditioning that global distribution on a particular subcommunity (e.g., CS), we can derive a *conditional* ranking of only those individuals within that subcommunity (e.g., Pr[page 1 | CS], Pr[page 2 | CS], etc.). Likewise, from the respective marginal probability of each subcommunity, we can infer what we call a *marginal* ranking[1] of subcommunities themselves (e.g., Pr[AI | CS], Pr[theory | CS], etc.). Conversely, it is straightforward to recover the global ranking by combining the conditional and marginal rankings using the chain rule. For example, Pr[page 1] = Pr[page 1 | AI] Pr[AI | CS] Pr[CS].

[1] Viewing each interior node as the root of a subtree, we informally refer to the ranking of the children of an interior node as a marginal ranking, although such a ranking is technically a *conditional* marginal ranking, conditioned on the subcommunity defined by that subtree.

Hence, to compute a global ranking of the individuals in an hierarchical social network, it suffices to compute marginal rankings at all interior nodes (i.e., rank the children of all interior nodes), and combine those marginal rankings via the chain rule. To facilitate recursive implementation, QuickRank localizes the computation of each marginal ranking: any links to or from leaves outside the subtree at hand are ignored in such computations. Beyond this computational motivation, localizing marginal ranking computations can be motivated by the following "peer-review principle:" *endorsements among peers (i.e., members of the same subcommunity) should be taken at face value, while other endorsements should be considered as only approximate*.

Intuitively, it is plausible that ranking information among individuals in a tightly-knit community would be more reliable than ranking information among individuals who are only loosely connected. Recall the citation index, a natural example of an hierarchical social network. When a researcher cites a topic in his area of expertise, he is likely to select the most appropriate references. In contrast, if for some reason a researcher with expertise in one area (e.g., computer science) is citing a result in another (e.g., sociology), he may choose only somewhat relevant references. Hence, we contend that the peer-review principle, which justifies localized marginal ranking computations, befits at least some application areas.

To fully implement the peer-review principle it is necessary to define some notion of approximate endorsements. To this end, we interpret an endorsement by an individual i in community A for another individual $j \neq i$ in another community $B \neq A$ as comprising part of an endorsement by A of B. More precisely, we aggregate endorsements by individuals in A for individuals in B into an endorsement by A of B by first scaling the endorsements from each i to each j by i's marginal rank, and then summing the resulting weighted endorsements. If we were to replace the target j of an endorsement by any other $j' \in B$, the resulting aggregate endorsement remains unchanged. In this sense, the original endorsement is viewed as "fuzzy" or "approximate." Moreover, by interpreting links originating at i as i's judgment, this aggregation process can be seen as an application of Bonacich's hypothesis (to obtain endorsements of each $j \in B$ by A) followed by a summation over all $j \in B$ (to obtain an endorsement of B).

Together, the principle of peer review and Bonacich's hypothesis lead to the QuickRank algorithm, which we illustrate on the example in Figure 9.1. We begin by restricting the

link graph to, say, the AI subdomain, thereby constructing a local link subgraph. Next, we apply any "flat" ranking scheme (e.g., degree and eigenvector centrality and PageRank) to this link subgraph to produce a marginal ranking of the pages in the AI subdomain (i.e., a distribution over 1 and 2). Then, we scale the links from 1 to 4 and 2 to 3 by the marginal ranks of 1 and 2, respectively, to generate links from AI to 4 and 3. Finally, we sum these results to produce an aggregate link from AI to theory.

Repeating this procedure for the theory and systems subdomains, we "collapse" each of the CS subdomains into a leaf, and substitute these subdomains for their corresponding web pages in the link graph. We then proceed recursively, constructing a local link subgraph, and computing a marginal ranking of the CS subdomains. Combining this marginal ranking with the marginal rankings of the web pages in each CS subdomain yields a single marginal ranking of all the web pages in the CS domain. We repeat this process until the entire hierarchy has been collapsed into a single node, at which point we obtain a ranking of all pages in the `edu.brown` domain.

We conclude this introduction by noting the following property of QuickRank: *The relative global ranking between two individuals is determined by their local ranks in the smallest community to which they both belong.* This property follows from the fact that scaling is the only operation which is performed on rankings (conditional rankings are scaled by marginal ranks), but scaling does not affect relative rankings.

Overview This paper purports to contribute to the literature on social network analysis by introducing the QuickRank algorithm. As suggested by the previous example, QuickRank is parameterized by a "BaseRank" procedure (i.e., a flat ranking scheme, such as degree centrality) used to compute marginal rankings. We begin in the next section by precisely defining BaseRank procedures and identifying desirable properties of such procedures. In Section 9.3, we present pseudocode for the QuickRank algorithm. We also consider to what extent QuickRank preserves our previously identified desirable properties of BaseRank procedures. Then, in Section 9.4, we provide sample QuickRank calculations. Our first example illustrates the distinction between stand alone "BaseRanks" and "QuickRanks," the rankings output by these schemes. A further example shows how QuickRank is potentially more resistant to link-spamming than corresponding BaseRank procedures. We conclude in Section 9.8. A discussion of related work is deferred to the QuickRank technical report, currently in preparation.

9.2 A Unified View of Flat Ranking Algorithms

QuickRank is parameterized by a flat (i.e., non-hierarchical) ranking algorithm, or a "BaseRank" procedure. In this section, we precisely define a BaseRank procedure, and we formulate the four flat ranking schemes mentioned in the introduction as such. We also present four desirable properties of BaseRank procedures, and discuss to what extent the four aforementioned ranking schemes satisfy these properties.

9.2.1 Preliminary Definitions

A social network encodes relationships among individuals in a society. Such a network can be represented by a *link graph*. Individuals $i, j \in \mathcal{I}$ are represented as *vertices*, and the fact that individual i relates to individual j is represented by a directed *link* from vertex i to vertex j, augmented by a nonnegative real-valued weight indicating the strength of i's relationship to j.

A *judgment* is a nonnegative, real-valued vector indexed on $\mathcal{I}$. We define an equivalence relation on judgments with r^1 and r^2 equivalent if $cr^1 = r^2$. For our purposes, a *ranking* is such an equivalence class $\langle r \rangle$ (although we often refer to a ranking by any representative of the class). A ranking has exactly one representative that is a probability distribution, which can be obtained by normalizing any other representative. Further, a ranking represents a consistent estimate of the relative merit of pairs of individuals: i.e., for all pairs of individuals i and j, the ranking of i relative to j, namely $\frac{r_i}{r_j} \in [0, \infty]$, is well-defined.

A *link graph* is a nonnegative, real-valued square matrix indexed on $\mathcal{I}$. We restrict attention to the case where the weights in the link graph may reasonably be interpreted as endorsements, rather than distances.[2] A *judgment* graph is a link graph further constrained to have *positive* diagonal entries. Each column in a judgment graph represents the judgment of one individual. The requirement that the diagonal be positive can be interpreted to mean that individuals are required to judge others relative to themselves. Whereas rankings are scale invariant, judgments are scale dependent.

In the introduction, we presented ranking schemes as operating on link graphs. That was a convenient oversimplification. More precisely, they map a judgment graph and a

[2] It is conceivable that QuickRank can be suitably modified to handle the distance interpretation by redefining the peer-review notion of approximation as aggregating by taking a minimum instead of summing, but we have not yet explored any applications of this sort.

prior ranking to a *posterior* ranking. We view the inference of a judgment graph from a link graph as a preprocessing step. This step might consist of inserting self-loops: replacing zeros on the diagonal with ones. In the case of the Web or a citation database, for example, such self-loops would model each web page or publication as implicitly referring to (i.e., endorsing) itself.

Analogously, we define a *BaseRank* procedure as a higher-order function that takes a judgment graph to a mapping which infers a posterior ranking from a prior. When used within the QuickRank algorithm, we require that the posterior ranking output by the BaseRank procedure be normalized to a probability distribution. The prior ranking may be viewed as the persuasion of the "center" (i.e., the implementer of the ranking scheme). A BaseRank procedure then is a means of aggregating the judgments of the individuals in the society, and the center, into a single collective posterior ranking.

Given a judgment graph R and a prior ranking $\langle r \rangle$, Bonacich's hypothesis suggests that we may infer a collective judgment as $r' = Rr$. In this way, individual j's posterior position is the sum of each individual i's conception of j, weighted by the prior rank of i. By ignoring scale in r', we can infer the posterior ranking $\langle r' \rangle$. Note that the result of these two inference steps is well-defined, in that $\langle r' \rangle$ depends only on $\langle r \rangle$ and not on r itself. We use the term *linear* to describe a BaseRank procedure whose mapping from a prior ranking to a posterior abides by Bonacich's hypothesis.

This inference rule shows up naturally in the case of two simple types of judgment graphs, namely, finite-state, discrete-time, stationary Markov processes and Bayesian updating. In the former case, the judgment graph corresponds directly to the probability transition matrix of the Markov process and the inference rule follows the corresponding reallocation of probability. In Bayesian updating, one is given a prior probability distribution $r_i = \Pr[A_i]$ over events A_i, together with the conditional probabilities $R_{ii} = \Pr[B \mid A_i]$ of some common event B. The Bayesian approach infers the posterior distribution $r'_i = \Pr[A_i \mid B]$ precisely as above: i.e., $r'_i = \frac{(Rr)_i}{\|Rr\|_1}$. In fact, any judgment graph can be expressed as the composite of these two types, a matrix with constant column sums and a diagonal matrix.

9.2.2 Sample BaseRank Procedures

We now describe how the four ranking schemes mentioned in the introduction (i.e., indegree, outdegree, eigenvector centrality and PageRank) can be viewed BaseRank procedures. We assume that the link graph has been pre-processed, with self-loops inserted as necessary, to yield an "initial" judgment graph. Since the inference step is fixed, the key step in a linear BaseRank procedure is the way in which a "final" judgment graph is inferred from the initial judgment graph. The degree centrality metrics and PageRank are examples of linear BaseRank procedures, as is eigenvector centrality under certain assumptions (see Theorem 9.2).

The indegree and outdegree of individual i are defined respectively, as follows: given an initial judgment graph R,

$$\text{IN}(i) = \sum_j R_{ij} \qquad \text{OUT}(i) = \sum_j R_{ji} \tag{9.1}$$

Both these centrality metrics can be understood as linear BaseRank procedures that infer a posterior ranking from a uniform prior. Indegree is simply the identity function: the initial and final judgment graphs are identical. Outdegree is the transpose operation: the initial and final judgment graphs are transposes of one another.

The PageRank algorithm is parameterized by a value $\epsilon \in (0, 1)$ and a distribution v, often referred to as a "personalization vector." In a preprocessing step, the columns of the judgment graph are normalized to yield a Markov matrix M. PageRank operates on the convex combination of M with the rank one Markov matrix vJ^t (where J ambiguously denotes any vector of all 1's), namely $M_\epsilon = (1-\epsilon)M + \epsilon v J^t$. This matrix is easily seen to be *unichain* (see Chapter 1), hence with a unique stable distribution v_∞. Moreover, Haveliwala and Kamvar (2003) have shown that M_ϵ has a second largest eigenvalue of $1-\epsilon$, so that $\lim_{k\to\infty} M_\epsilon^k v_0 = v_\infty$, for any initial distribution v_0, with convergence as $(1-\epsilon)^k$. This result follows alternatively by writing v_∞ as the limit of a geometric series:

Theorem 9.1. *If M is a Markov matrix and $M_\epsilon = (1-\epsilon)M + \epsilon v J^t$, then*

$$v_\infty = \lim_{k\to\infty} M_\epsilon^k v_0 = \epsilon \sum_{i=0}^{\infty} (1-\epsilon)^i M^i v \tag{9.2}$$

This theorem implies that PageRank is a linear BaseRank procedure, which takes an initial judgment graph M to a final judgment graph $\epsilon \sum_{i=0}^{\infty} (1-\epsilon)^i M^i$. The prior ranking

corresponds to the personalization vector and the posterior ranking is a discounted sum of all the inferred rankings (including the prior).

Unlike degree centrality and PageRank, which we have shown are linear BaseRank procedures, eigenvector centrality is not. Given a judgment graph R and an prior ranking v_0, the algorithm infers a sequence of posterior rankings $v_{n+1} = \frac{Rv_n}{\|Rv_n\|_1}$. It can be shown that this sequence eventually converges to a fixed point v_∞, which can be interpreted as the collective ranking. Moreover, this iterative process can be expressed as a linear inference $v_\infty = \frac{R_\alpha v_0}{\|R_\alpha v_0\|_1}$, where α, and hence R_α, depend on the support of v_0. In particular, eigenvector centrality is a *piecewise*-linear BaseRank procedure. In the special case where the judgment graph is strongly-connected (i.e., R is irreducible), eigenvector centrality is linear, because R_α is constant (i.e., independent of α) and v_∞ is independent of v_0. Formally,

Theorem 9.2. *If a judgment graph $R \geq 0$ is irreducible with non-zero diagonal, there exists a unique ranking $v > 0$, such that $\|v\|_1 = 1$ and $Rv = \rho(R)v$, where $\rho(R)$ is the magnitude of the largest eigenvalue of R. Moreover, for any $v_0 \geq 0$, if $v_{n+1} = \frac{Rv_n}{\|Rv_n\|_1}$, $\lim_{n\to\infty} v_n = v$. That is, $v_\infty = v$ and for all α, $R_\alpha = vJ^t$.*

9.2.3 Generalized Proxy Voting

If we view each individual's rank as a collection of proxy (i.e., infinitely divisible and transferable) votes, then a judgment graph may be interpreted as a *proxy-vote specification* indicating how each individual is willing to assign his proxy votes to others. Given a prior ranking (i.e., an initial allocation of proxy votes), the posterior inferred by a linear BaseRank procedure is a reallocation based on the results of a single round of proxy voting. More generally, in *generalized proxy-voting* (GPV), individuals cast their votes repeatedly over time (i.e., each posterior serves as a prior in the next round), until ultimately, the sequence of posteriors is averaged into a final vote count: i.e., a final ranking.

While historically PageRank has been viewed in terms of a "random-surfer" model (cf. Page et al. (1998)), Theorem 9.1 suggests that it may be more aptly viewed as a GPV mechanism with a discount factor $\gamma \in [0, 1)$. In particular, for a given prior ranking v, the posterior computed by PageRank can be expressed as $(1-\gamma)^{-1}\sum_{i=0}^{\infty}\gamma^i M^i v$. Notice that this is just the average of the inferred rankings $M^i v$, where i is distributed geometrically with mean γ. It is natural to generalize to allow weighting by arbitrary distributions,

$\sum_{i=0}^{\infty} \alpha_i M^i v$, or even as the limit of such, $\lim_{N\to\infty} \sum_{i=0}^{N} \alpha_{i,N} M^i v$. Formally, we define a generalized proxy-voting mechanism as a (linear) BaseRank procedure that takes an initial judgment graph M into a final judgment graph $\lim_{N\to\infty} \sum_{i=0}^{N} \alpha_{i,N} M^i$.

Observe that all the flat ranking schemes mentioned above, except outdegree, are not only linear BaseRank procedures, but can be seen as GPV mechanisms as well. Indegree is a trivial instance of GPV with $\alpha_{i,N} = \delta_{i,1}$. By Theorem 9.1, PageRank is a GPV mechanism with $\alpha_{i,N} = \epsilon(1-\epsilon)^i$. Finally, if we restrict attention to irreducible judgment graphs, eigenvector centrality is a GPV mechanism, with $\alpha_{i,N} = \begin{cases} \frac{1}{N+1} & \text{if } 0 \leq i \leq N \\ 0 & \text{otherwise} \end{cases}$. This final claim follows Theorem 9.2 and the well-known fact that $\lim_{i\to\infty} s_i = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} s_i$. Although outdegree, which takes R to R^t is linear, it is not a GPV mechanism.

9.2.4 Axioms

Next, we identify two types of judgment graphs that have natural interpretations, and on which a particular behavior for a BaseRank procedure seems preferred. First, consider the identity matrix I as a judgment graph—the *identity* graph—in which each individual ranks himself infinitely superior to all others. Such a ranking graph provides no basis for modifying a prior ranking. Thus, on this input, it seems reasonable that a BaseRank procedure should act as the identity function (i.e., posterior = prior).

Second, consider the case of a *consensus* graph, that is, a judgment graph xy^t, where x is a distribution and y_i is individual i's arbitrary scaling factor. In other words, a consensus graph is a rank 1 matrix: everyone agrees on the ranking x, up to a multiple. Since there is consensus among the individuals in the society, we contend that any prior ranking should be ignored. A BaseRank procedure should simply return the consensus x. We restate these two properties succinctly, as follows:

Identity: $BaseRank(I) = \text{id}$

Consensus: $BaseRank(xy^t) = x$

Another important issue associated with ranking schemes is that of manipulation via "link spamming." The goal of link spamming is to game a ranking system by creating many false nodes, sometimes called sybils (Cheng and Friedman, 2006), that link to some

node n, thereby attempting to influence the rank of node n. Web spamming is a particularly popular form of link spamming (Gyongyi and Garcia-Molina, 2004).

A judgment graph inhabited by sybils takes the following form: $M' = \begin{bmatrix} M & \overline{N} \\ 0 & \overline{M} \end{bmatrix}$, where M is the original judgment graph (i.e., without the sybils), $\overline{N}$ describes the links from the sybils to existing members of the society, and $\overline{M}$ describes the links among sybils. Since sybils are new to the community, and hence unknown its original members, we assume that there are no links from those members to sybils.

Observe that generalized proxy-voting mechanisms are spam-resistant in the following sense: Given a prior ranking which places no weight on sybils, the posterior ranking computed with respect to the modified judgment graph M' is, for all intents and purposes, equivalent to the posterior ranking computed with respect to the original judgment graph M. That is,

Theorem 9.3. *If* $M' = \begin{bmatrix} M & \overline{N} \\ 0 & \overline{M} \end{bmatrix}$, $v' = \begin{bmatrix} v \\ 0 \end{bmatrix}$, *and* $BaseRank(\cdot) = \lim_{N\to\infty} \sum_{i=0}^{N} \alpha_{i,N} (\cdot)^i$, *then* $BaseRank(M')v' = \begin{bmatrix} BaseRank(M)v \\ 0 \end{bmatrix}$.

For example, since PageRank is a GPV mechanism, we apply Theorem 9.3 to show that the posterior ranking of non-sybils is unaffected by their presence, if we assign sybils a prior rank of 0. In other words, if sybils can be detected *a priori*, then PageRank may be rendered immune to such an attack. Although the corresponding Markov matrix need not be irreducible for such a "personalization" vector, we conclude from Theorem 9.1 that the Markov process converges for *all* prior rankings v_0. Note that this conclusion follows specifically from our interpretation of PageRank as a GPV mechanism, as opposed to the traditional "random surfer" model.

Table 9.1 summarizes how each of the four ranking schemes discussed in this section behave with respect to the four properties of BaseRank procedures discussed in this section. PageRank does *not* satisfy the consensus property because it is always biased to some degree by the prior ranking. However, using the notation introduced above, if we instead define $M_\epsilon = (1-\epsilon)M + \epsilon M v J^t$, the resulting algorithm satisfies all four properties. This modified PageRank corresponds to a linear BaseRank procedure with final judgment graph

Property	Indegree	Outdegree	Eigenvector	PageRank
Linear	Yes	Yes	*No*	Yes
GPV	Yes	*No*	Yes	Yes
Identity	Yes	Yes	Yes	Yes
Consensus	Yes	Yes	Yes	*No*

Table 9.1: Some properties of ranking schemes.

$\epsilon \sum_{i=0}^{\infty} (1-\epsilon)^i M^{i+1}$, that is, the posterior is a discounted sum of all inferred rankings *excluding* the prior.

Fundamentally, QuickRank's design is based on the two key ideas discussed in the introduction, namely the peer-review principle and Bonacich's hypothesis. However, as QuickRank is parameterized by a BaseRank procedure, it is also designed to preserve the Identity and Consensus properties. In the next section, we detail the algorithm and argue informally that it indeed preserves these two properties of BaseRank procedures, although it fails to preserve linearity. When we present sample calculations in Section 9.4, we note that QuickRank preserves the spam-resistance of its BaseRank procedure, and we illustrate its potential to resist spam even further.

9.3 QuickRank: The Algorithm

QuickRank operates on a hierarchical social network, that is a judgment[3] graph R whose vertices are simultaneously leaves of a tree T. At a high level, QuickRank first ranks the leaves using the link information contained in the local subgraphs; it then propagates those local[4] rankings up the tree, aggregating them at each level, until they have been aggregated into a single global ranking. Ultimately, *a node's QuickRank is the product of its own local rank and the local rank of each of its ancestors.* QuickRank is parameterized by a BaseRank procedure, which it uses to compute local rankings. It also takes as input a prior ranking of the leaves. It outputs a posterior distribution.

Although we present QuickRank pseudo-code (see Algorithm 5) that is top-down and recursive, like many algorithms that operate on trees, the simplest way to visualize the

[3] As above, we assume the link graph has been pre-processed to form a judgment graph.

[4] Whereas in the introduction, we used the term marginal, we now use the term local to refer to the ranking of a node's children. The salient point here is: this ranking is computed using strictly local information.

QuickRank algorithm is bottom-up. From this point of view, QuickRank repeatedly identifies "collapsible" nodes in T, meaning the root nodes of subtrees of depth 1, and collapses them into leaf nodes (i.e., subtrees of depth 0) until there are no further opportunities for collapsing: i.e., until T itself is a leaf node. Collapsing node n entails: (i) computing a local ranking at n, that is a ranking of n's children, and (ii) based on this local ranking, aggregating the rankings and the judgments of n's children into a single ranking and a single judgment, both of which are associated with n.

Note that QuickRank is a well-defined algorithm: that is, the order in which local rankings are computed does not impact the global ranking. This property is immediate, since QuickRank propagates strictly local calculations up the tree in computing its global output. Moreover, the collapse operation replaces a subtree of depth 1 with a subtree of depth 0 so that QuickRank is guaranteed to terminate.

Data Structures Algorithm 5 takes as input T_n, subtree of T rooted at node n, and returns two data structures: (i) a ranking of all leaves (with support only on T_n) and (ii) a judgment, which is the average of all judgments of T_n's leaves, weighted by the ranking computed in (i). At leaf node n, the ranking is simply the probability distribution with all weight on n, denoted e_n, and the judgment is given by R_n.

Computing Local Rankings Recall that the main idea underlying QuickRank is to first compute local rankings, and to then aggregate those local rankings into a single global ranking. Given a collapsible node n, a local ranking is a ranking of n's children. To compute such a ranking, QuickRank relies on a BaseRank procedure.

There are two inputs to this BaseRank procedure. The first is n's local (i.e., marginal) prior ranking. The second is a local judgment graph M. For j and k both children of node n, the entry of M in the row corresponding to k and the column corresponding to j is the aggregation of all endorsements from leaves in T_j to leaves in T_k, equal to the sum of all entries in the jth judgment corresponding to leaves of T_k.

Aggregating Rankings and Links To aggregate the rankings of n's m children into a single ranking associated with n, QuickRank averages the rankings $r^1, \ldots, r^m$ according to the weights specified by the local ranking r. If we concatenate the m rankings into a matrix $Q = \left[\begin{array}{ccc} r^1 & \cdots & r^m \end{array} \right]$, then the aggregation of rankings can be expressed simply

as Qr. Also associated with each child j of a collapsible node n is a judgment l^j. These judgments are aggregated in precisely the same way as rankings.

Algorithm 5 QuickRank(node n)

if (n.isLeaf())
    return($\langle n$.getJudgment(), $e_n\rangle$);
m = n.numChildren()
for ($j = 1$ to m) {
    $\langle l^j, r^j\rangle \leftarrow$ QuickRank(n.getChild(j))
    for ($k = 1$ to m) {
        M_{kj} = Sum(l^j, n.getChild(k))
    }
}
$P = \left[\begin{array}{ccc} l^1 & \ldots & l^m \end{array}\right]$
$Q = \left[\begin{array}{ccc} r^1 & \ldots & r^m \end{array}\right]$
r = BaseRank(M, n.getLocalPriorRanking())
return($\langle Pr, Qr\rangle$);

We now argue that if the BaseRank procedure satisfies the Identity and Consensus properties, then so, too, does QuickRank. First, notice that, when restricted to any subcommunity (i.e., square, diagonal block), an identity or consensus graph yields the same type of graph again. Moreover, aggregating links in such a community within the original graph (i.e., summing rows and averaging columns) also results in the same type of graph. Consequently, if QuickRank employs a BaseRank procedure with the Identity property, it will output the prior distribution on the identity graph, since the prior local rankings will remain unchanged at each level in the hierarchy.

Now consider a consensus graph with ranking x s. t. $\|x\|_1 = 1$. Restriction to a subcommunity gives a consensus graph on the corresponding conditional distribution of x. Likewise, aggregation produces a consensus graph on the corresponding marginal distribution of x. If QuickRank employs a BaseRank algorithm with the consensus property on a consensus graph, it will gradually replace the prior distribution at the leaves with the conditional distributions of x, until it finally outputs x itself.

We conclude this section by pointing out that, even if the BaseRank procedure is linear, QuickRank may not be expressible as a linear inference. Normalizing local rankings to form distributions can introduce non-linearities. In the next section, we provide sample QuickRank calculations.

9.4 Examples

We now present two examples that verify our intuition regarding QuickRank and illustrate some of its novel features. Recall that QuickRank, as it operates on an hierarchical social network (HSN), is parameterized by a prior ranking and a BaseRank procedure.

First, consider the HSN shown in Figure 9.2a. The hierarchy is drawn using solid lines. The link graph is indicated by dotted lines between the numbered leaves. All weights are assumed to be 1. Computing QuickRanks for this HSN, varying the BaseRank procedure among indegree, eigenvector centrality, and PageRank,[5] but always assuming a uniform prior ranking, leads to the rankings, cardinal and ordinal, shown in Table 9.2. The values in the posterior distributions have been rounded; hence, the ordinal rankings more precisely reflect the exact values in those distributions.

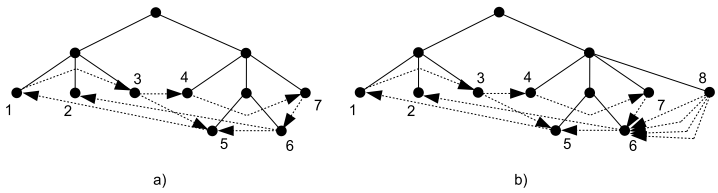


Figure 9.2: Two examples of hierarchical social networks.

Table 9.2: BaseRanks and QuickRanks from Figure 9.2a and uniform prior.

		Indegree	Eigenvector	PageRank
Flat	cardinal	{0.13, 0.13, 0.13, 0.13, 0.2, 0.13, 0.13}	{0.19, 0.08, 0.16, 0.14, 0.22, 0.10, 0.12}	{0.14, 0.32, 0.11, 0.09, 0.14, 0.09, 0.11}
	ordinal	$5 > 1 = 2 = 3 = 4 = 6 = 7$	$5 > 1 > 3 > 4 > 7 > 6 > 2$	$2 > 1 > 5 > 3 > 7 > 6 > 4$
QuickRank	cardinal	{0.10, 0.10, 0.19, 0.09, 0.23, 0.11, 0.18}	{0, 0, 0.41, 0, 0.59, 0, 0}	{0.04, 0.14, 0.25, 0.04, 0.41, 0.06, 0.06}
	ordinal	$5 > 3 > 7 > 6 > 1 = 2 > 4$	$5 > 3 > 1 = 2 = 4 = 6 = 7$	$5 > 3 > 2 > 7 > 6 > 1 > 4$

For each BaseRank procedure, we list two pairs of rankings: that which results from ignoring the hierarchy, and that which results from exploiting it using QuickRank. When we ignore the hierarchy, all three algorithms rank leaf 1 above (or equal to) 3. However, since 1 defers to 3 (i.e., 1 endorses 3, but not vice versa), based on our peer-review principle,

[5]The results of ranking with outdegree are not qualitatively different, but are omitted for lack of space.

3 should be ranked higher than 1. This outcome indeed prevails in the QuickRanks, for all three BaseRank procedures.

When using a uniform prior ranking, the resulting rankings are not biased by the depth at which individuals reside in the hierarchy. If such a bias is desirable, however, it can be easily achieved with a non-uniform prior. For example, a prior ranking of $\frac{1}{12}\{2, 2, 2, 2, 1, 1, 2\}$ with indegree as BaseRank yields a posterior ranking of $\{.10, 0.10, 0.19, 0.12, 0.18, 0.09, 0.23\}$, which corresponds to an ordinal ranking of $7 > 3 > 5 > 4 > 1 = 2 > 6$. Whereas 5 was ranked higher than 7 with a uniform prior, 7 ranks highest with this biased prior.

As an added benefit, QuickRank may be more resistant to link spamming than BaseRank procedures that do not exploit hierarchies. To demonstrate this phenomenon, in Figure 9.2b, we introduce a sybil, leaf 8, into our original example to try and raise the rank of 6 by recommending it highly. Note the multiplicity of links from 8 to 6.

Table 9.3: Fig. 9.2b with Indegree as BaseRank.

		Uniform Prior	Weighted Prior
Flat	cardinal	{0.10, 0.10, 0.10, 0.10, 0.15, 0.30, 0.10, 0.05}	{0.13, 0.13, 0.13, 0.13, 0.2, 0.13, 0.13, 0.0}
	ordinal	$6 > 5 > 1 = 2 = 3 = 4 = 7 > 8$	$5 > 1 = 2 = 3 = 4 = 6 = 7 > 8$
QuickRank	cardinal	{0.09, 0.09, 0.18, 0.06, 0.28, 0.14,0.11, 0.06}	{0.10, 0.10, 0.19, 0.09, 0.23, 0.11, 0.19, 0.0}
	ordinal	$5 > 3 > 6 > 7 > 1 = 2 > 4 = 8$	$5 > 3 > 7 > 6 > 1 = 2 > 4 > 8$

Applying QuickRank with indegree as BaseRank to this example yields the rankings shown in Table 9.3. Using a uniform prior, the sybil is able to raise the rank of 6 over 7 and 6 over 4, whether we exploit the hierarchy (i.e., use QuickRank) or not (i.e., compute indegrees directly). QuickRank cannot prevent this outcome, since the sybil is an accepted member of 4's and 7's community. However, the influence of the sybil is somewhat mitigated under QuickRank. Since the resulting ranking must respect the hierarchy, the effect of the sybil is to raise the ranks of *both* 5 and 6 (i.e., both values in the posterior distribution). No amount of link spam from a sybil outside their local community can increase the rank of 6 relative to 5.

Moreover, if one is able to identify sybils *a priori*, by setting the prior ranks of sybils to zero, one can reduce their influence even further. If we use a prior ranking which is weighted against the sybil, say uniform over 1-7 and zero on 8, Table 9.3 shows that indegree produces the same rankings as in Table 9.2, that is, *without* the sybil, whether we

exploit the hierarchy or not. In general, Theorem 9.3 states that any BaseRank procedure which is a GPV mechanism will necessarily exhibit this same behavior. QuickRank is not a GPV scheme (recall that QuickRank is nonlinear but that GPV schemes are linear). Still, QuickRank preserves the spam-resistance property characteristic of GPV mechanisms.

9.5 Experiments

In this section, we discuss some preliminary experiments we performed to validate our **QuickRank** technique. Specifically, we compare the performance of **QuickRank** utilizing two different BaseRank algorithms (indegree and PageRank) on three sample information retrieval tasks, the 2002, 2003, and 2004 TREC Topic Distillation Tasks, part of the annual TREC competition[6]. As described in the 2003 report, "the topic distillation task involves finding relevant homepages, given a broad query," where "a good homepage [corresponds] to a site which:

- Is principally devoted to the topic,
- Provides credible information on the topic, and
- Is not part of a larger site also principally devoted to the topic." (Craswell and Hawking, 2003).

Queries were applied to a corpus of U.S. government web pages, the .GOV test collection, containing about 1.25 million pages.[7] In each of 2002 and 2003, the task involved 50 queries, while in 2004 it used 75 queries. For each query, the TREC organizers compiled a list of pages which it deemed as sufficiently good responses, or "query-relevant", for the task. Specifically, the pages returned by all the entrants to the competition were rated by human judges. Those with sufficiently high scores were deemed query-relevant, and a corresponding list of "qrels" were then published by the TREC organizers for future research. Notice that this set of qrels is thus biased to favor competitors against non-entrants of the competition, such as our **QuickRank** implementation.

[6] `http://trec.nist.gov/`

[7] `http://ir.dcs.gla.ac.uk/test_collections/`

Apart from the third criteria, the goal of the Topic Distilation (TD) task is to strike a balance between relevance and "authoritativeness" (authority) in Web search. While we believe that **QuickRank** should provide a meaningful (query-independent) measure of the authority of a web page (as judged by the community of web page publishers), we needed an additional (query-dependent) technique to filter web pages for relevance to the given query. We used Apache Lucene, a "high-performance, full-featured text search engine library written entirely in Java."[8] We then took a convex combination of the resulting rank scores from each technique, with mixing parameter, α, to obtain the ranking of each page of a query.

In order to apply **QuickRank**, we needed to infer a hierarchical, social network on the corpus of web pages. We used the (unweighted) link graph which the TREC organizers distributed with the corpus, converting it to a judgment graph as described in section 9.2.2. We used the URL hierarchy as described in section 9.2.2, except for efficiency, we collapsed subtrees to insure that there were a minimum 200 leaves per node and the hierarchy had a maximum depth of k, where $k = 0, \ldots, 7$. This is clearly not the most informative hierarchy; it was simply the most readily available one. Thus, in our results we focus attention to compare depth 0 (i.e., simply applying the BaseRank algorithm) and depth 1. The branching factor at depth 1 is roughly 250

So that we could compare our results with those of the TREC competitors, we applied several standard measures to the ranking resulting produced each query. If, for a given query, H represents the set of all documents retrieved and $Qrels$ is the set of documents judged to be query-relevant, while H_n is the subset of n top-ranked documents, we may define the following measures on the retrieval system:[9]

- *Success at* n: $S@n = [H_n \cap Qrels \neq \emptyset]$
- *Recall at* n: $R@n = \frac{|H_n \cap Qrels|}{|Qrels|}$
- *Precision at* n: $P@n = \frac{|H_n \cap Qrels|}{|H_n|}$
- *Precision at* R: $P@R = \frac{\left|H_{|Qrels|} \cap Qrels\right|}{|Qrels|}$
- *Average Precision*: $AP = \sum_{n=1}^{|H|} P@n \frac{\left[H_n - H_{n-1} \subset Qrels\right]}{|Qrels|}$

[8] `http://lucene.apache.org/java/docs/`

[9] While $|H_n| = n$, we write it out to show the symmetry in the definitions of precision and recall.

Notice that these measures assume that H may be ordered *without ties*. Since **QuickRank** makes so such guarantees, these measures may have been unduly affected by the order in which equally ranked pages were processed. The same set of measures were not applied uniformly across all three years; we report only those statistics reported from each year.

We give the results of our experiments in Tables 9.4-9.6. Each column represents the average of the given measure over all queries for that year's competition. The rows correspond to the three selected competitors, along with five runs of our algorithm with varying parameters. We chose to include the top- and bottom-scoring competitors for each year, along with a third competitor which has roughly the median score for reference. Likewise, we include the results with $\alpha = 1$, which we label "Lucene", since the ranking of the query results is entirely determined by Lucene's relevance score.

Table 9.4: Comparison with TREC 2002 competitors

P@10	α	**Algorithm**	**Depth**
0.251	-	thutd5	-
0.198	0.99	PageRank	0
0.194	-	mu525	-
0.190	0.99	Indegree	1
0.190	0.99	Indegree	0
0.184	0.99	PageRank	1
0.182	1	Lucene	-
0.057	-	ajouai0210	-

For each query, the rank scores produced by Lucene and **QuickRank** were on very different scales. The latter tends to be more exponential, while the precise nature of the former is unclear. In addition, $|H| \gg |Qrels|$. Thus, how to combine the two scores was problematic. We first converted each rank score to a linear, 0 to 1 scale; that is, the scores of consecutively ranked documents were $\frac{1}{|Qrels|}$ and $\frac{1}{|H|}$ apart, respectively. We then used a simple convex combination, with weight parameter α, to combine the results. Again, since it is unclear how much weight to assign to topic relevance versus authorithy, we ran our experiments for various values of $alpha$ to discover a proper value, which ranged from $.95$ to $.99$ over the three years.

Along with the three reference competitors and the results of simply using relevance score alone (i.e., Lucene), we report the performance of **QuickRank** using indegree and

Table 9.5: Comparison with TREC 2003 competitors

P@10	P@R	AP	α	Algorithm	Depth
0.124	0.164	0.154	-	csiro03td03	-
0.090	0.114	0.099	0.97	Indegree	1
0.086	0.105	0.097	0.97	Indegree	0
0.082	0.086	0.089	1.00	Lucene	-
0.074	0.092	0.088	0.97	PageRank	0
0.062	0.078	0.087	0.97	PageRank	1
0.092	0.092	0.070	-	meijihilw1	-
0.032	0.028	0.023	-	C2B	-

PageRank as BaseRank algorithms. When depth is 0, these techniques are just the standard algorithms (i.e., without exploiting the hierarchy). From the results in Tables 9.4-9.6, we can see that indegree at depth 1 generally performed well, and in particular, it always performed indegree at depth 0. Moreover, it almost always outperformed PageRank at either depth.

This suggests a number of practical benefits to **QuickRank**. Remember that PageRank was designed to mitigate the manipulability of indegree via link-spamming. However, from these experiments and our discussion in Section 9.4, we see that simply by applying **QuickRank** with indegree at depth 1, we can limit the influence of link-spamming without sacrificing the quality of our resulting rankings. Notice that this is even more striking, since indegree is much simpler and faster to compute than PageRank.

Table 9.6: Comparison with TREC 2004 competitors

S@1	S@5	S@10	P@10	R@1000	AP	α	Algorithm	Depth
0.507	0.773	0.893	0.249	0.777	0.179	-	uogWebCAU150	-
0.213	0.680	0.773	0.151	0.590	0.123	0.95	Indegree	1
0.253	0.680	0.813	0.163	0.590	0.120	0.95	Indegree	0
0.333	0.64	0.76	0.199	0.647	0.115	-	MU04web1	-
0.227	0.587	0.707	0.135	0.586	0.093	0.95	PageRank	0
0.080	0.400	0.573	0.109	0.569	0.075	1.00	Lucene	-
0.187	0.533	0.600	0.097	0.582	0.074	0.95	PageRank	1
0.067	0.147	0.173	0.029	0.147	0.018	-	irttil	-

9.6 Discussion: Implicit Hierarchical Structure

Some networks may come equipped with an explicit hierarchical structure (e.g., the Web's URL tree), but others may not. For networks in the latter category, it has been argued that many (social) networks tend to exhibit hierarchical structure at least implicitly (Simon, 1962). To run QuickRank on such a network, it would be necessary to infer this hierarchical structure. Even for networks in the former category, it may be worthwhile to infer an alternative hierarchical structure. In the case of the Web for example, QuickRanks may be more useful if pages are categorized into a topic hierarchy, rather than according to the URL tree.

It is possible to imagine a number of ways to infer an implicit hierarchical structure, given a network whose nodes are documents (e.g., Web pages, email messages, or publications). On the one hand, one could rely solely on the textual content of the documents (Blei et al., 2004). On the other hand, one could rely solely on the underlying graph-theoretic structure. In the case of the Web, it has been observed that the URL tree is reflective of the hierarchy that would be inferred based on its graph-theoretic structure (Eiron and McCurley, 2004). In principle, one could also rely on some combination of both approaches.

A difficulty arises in that some nodes in a network may not fit squarely in one category. For example, Arnold Schwarzenneger could be classified as both an actor and a politician. Alternatively, an algorithm that infers an implicit hierarchical structure may output a probability that each node belongs to each category. For example, Arnold Schwarzenneger could be classified as an actor with probability 0.9 and a politician with probability 0.1. We are developing natural extensions of the basic QuickRank algorithm that operate on hierarchical structures like these.

9.7 Related Work

The idea of constructing a global ranking by combining combining local rankings is not new. Indeed the electoral college is based on the same basic principle. Each state holds a local presidential election, the global outcome of which is determined by weighing the local outcomes according to the importance—in this case, the size of the Congressional

delegation—of each state.[10]

More to the point, Kamvar et al. (2003a) apply similar methodology to rank web pages in their algorithm, BlockRank, which is designed to exploit the block structure they observe in the Web. They do not recursively apply their reasoning, however. They combine domain and subdomain rankings only once, and then initialize PageRank with the resulting distribution, in an attempt to speed up the usual PageRank computation.

9.8 Conclusion

Social network, or link, analysis is regularly applied to information networks to compute rankings (Garfield, 1972; Kleinberg, 1998; Page and Brin, 1998; Page et al., 1998) and to social networks (Bonacich, 1972; Hubbell, 1965; Katz, 1953; Wasserman and Faust, 1994) to determine standing. We discuss two examples of information networks with inherent hierarchical structure: the Web and citation indices. Social networks, like the Enron email database, also exhibit hierarchical structure. Simon (1962) suggests that such hierarchies are ubiquitous:

> Almost all societies have elementary units called families, which may be grouped into villages or tribes, and these into larger groupings, and so on. If we make a chart of social interactions, of who talks to whom, the clusters of dense interaction in the chart will identify a rather well-defined hierarchic[11] structure.

Still, to our knowledge, link analysis procedures largely ignore any hierarchical structure accompanying an information or social network. In this paper, we introduced QuickRank, a link analysis technique for ranking individuals that exploits hierarchical structure. The foundational basis for QuickRank is the peer-review principle, which implies that the relative ranking between two individuals be determined by their local ranks in the smallest community to which they both belong. This principle, together with an hypothesis due

[10] QuickRank, applied to presidential elections, would normalize the popular vote in each state, and then weigh the resulting distributions by the corresponding number of electoral votes, a process which reduces to plurality voting.

[11] Simon's use of the terminology "hierarchic" is slightly broader than our use of "hierarchical structure," by which we mean tree structure. Still, the point remains: hierarchies (or approximations thereof) arise naturally in societies.

to Bonacich, leads to a recursive algorithm which is scalable, parallelizable, and easily updateable.

For a large-scale network such as the Web, we anticipate that QuickRank will yield substantial computational gains over standard ranking methods (e.g., calculating PageRanks via the power method). Moreover, it appears more resistant to link spamming than other popular ranking algorithms on contrived examples, although it remains to verify this claim empirically.

In ongoing research, we are attempting to empirically validate the merits of QuickRanks computed with some BaseRank procedure as compared to the ranking computed by the BaseRank procedure itself. Specifically we are augmenting Lucene, an open source Web search engine, with QuickRanks, PageRanks, and indegree ranks in order to measure the precision and recall of the augmented tool on the topic distillation queries from the TREC 2002, 2003, and 2004 web tracks.

Appendix A

Review of Linear Algebra

Throughout this thesis, we assume basic knowledge of vector spaces. Here we remind the reader of many of those specifics on which we heavily rely.

For any matrix, M, the *kernel* of M, sometimes called the *nullspace* of M, is defined as follows: $\ker M = \{v \in \mathbb{R}^n \mid Mv = 0\}$. Likewise, here is the definition of the *image* of M: $\operatorname{im} M = \{Mv \mid v \in \mathbb{R}^n\}$. The *span* of a set of vectors is the set of all linear combinations of those vectors. The image of M is sometimes called the *columnspace* of M because it is the span of the columns of M.

A finite set of vectors $\{v_i \in V \mid 1 \leq i \leq k\}$ is said to be *linearly independent* iff $\alpha_i = 0$, for all $1 \leq i \leq k$, whenever $\sum_{i=1}^{k} \alpha_i v_i = 0$, i.e., 0 cannot be expressed as a non-trivial linear combination of the vectors in the set. A *basis* for a vector space, V, is a linearly independent set of vectors whose span is V. The *dimension* of V is the cardinality of any basis (all bases have the same cardinality).

With these definitions in hand, we now state without proof two important theorems from linear algebra.

Theorem A.1. *For any $m \times n$ matrix, M,*

a) $\dim \operatorname{im} M = \dim \operatorname{im} M^t$. *We call this value the* rank *of* M.

b) $\operatorname{rk} M + \dim \ker M = n$, *and* $\operatorname{rk} M + \dim \ker M^t = m$.

c) *When M is square (when $m = n$),* $\dim \ker M = \dim \ker M^t$. *We call this value the* nullity *of* M.

Theorem A.2. *For any $m \times n$ matrix, M,*

a) If $LM = I$ for some $n \times m$ matrix L, we call M left-invertible *with* left-inverse L. *M is* left-invertible *iff M is* injective *iff* $\ker M = 0$.

b) If $MR = I$ for some $n \times m$ matrix R, we call M right-invertible *with* right-inverse R. *M is* right-invertible *iff M is* surjective *iff* $\ker M^t = 0$.

c) If M is square, then M is surjective iff M is injective iff M is invertible *with* inverse M^{-1} *such that $M^{-1}M = MM^{-1} = I$.*

The addition of two matrices is well-defined iff both matrices have the same dimensions. The multiplication MM' of an $m \times n$ matrix M and an $m' \times n'$ matrix M' is well-defined iff $n = m'$, and the resulting matrix will be of dimension $m \times n'$. Here are some simple observations about how the kernel of a matrix behaves with respect to matrix multiplication and addition.

Lemma A.3. *For any matrices, A and B, such that AB and BA are well-defined (i.e., if A is $m \times n$, B is $n \times m$),*

a) $\ker A \cap \operatorname{im} B = B \ker AB$

b) if B is surjective, then $\ker A = B \ker AB$*;*

c) if B is injective, then $\ker A = \ker BA$.

For any matrices, C and D, such that $C + D$ is well-defined (i.e., C and D have the same dimension),

d) if $\operatorname{im} C \cap \operatorname{im} D = 0$, *then* $\ker(C + D) = \ker C \cap \ker D$.

Proof. Proof of part a): If $v \in \ker AB$, then $A(Bv) = (AB)v = 0$, so that $Bv \in \ker A \cap \operatorname{im} B$, i.e., $B \ker AB \subset \ker A \cap \operatorname{im} B$. Conversely, any $w \in \ker A \cap \operatorname{im} B$ may be written as $w = Bv$ for some v. Since $ABv = Aw = 0$, $v \in \ker AB$, and $w = Bv \in B \ker AB$, so that $\ker A \cap \operatorname{im} B \subset B \ker AB$. Therefore, $\ker A \cap \operatorname{im} B = B \ker AB$.

Proof of part b): Further, if B is surjective, then $\ker A \subset \operatorname{im} B$, so that $\ker A = B \ker AB$.

Proof of part c): Now assume that B is injective, i.e., if $Bv = a$ and $Bw = a$, then $v = w$. If $w \in \ker BA$, then $B(Aw) = (BA)w = 0$. Since B is injective, $\ker B = 0$, and so $Aw = 0$, i.e., $w \in \ker A$. Therefore, $\ker BA \subset \ker A$. Likewise, if $v \in \ker A$, then $B(Av) = B0 = 0$, so that $v \in \ker BA$, $\ker A \subset \ker AB$, and $\ker A = \ker BA$.

Proof of part d): Finally, if $v \in \ker C \cap \ker D$, then $(C+D)v = Cv + Dv = 0 + 0 = 0$, so that $v \in \ker(C + D)$. Conversely, assume that $\operatorname{im} C \cap \operatorname{im} D = 0$. Now, if $v \in \ker(C + D)$, then $Cv + Dv = (C + D)v = 0$, so that $Cv = -Dv = D(-v) \in \operatorname{im} C \cap \operatorname{im} D$. Ty assumption, this is 0, so $Cv = Dv = 0$. Therefore, $v \in \ker C \cap \ker D$, and $\ker(C + D) = \ker C \cap \ker D$.

□

Bibliography

Nir Ailon. Aggregation of partial rankings, p-ratings and top-m lists. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 415–424, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

David M. Blei, Thomas L. Griffiths, Joshua B. Tenenbaum, and Michael I. Jordan. Hierarchical topic models and the nested chinese restaurant process. In Sebastian Thrun, Lawrence K. Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, pages 17–24, Cambridge, MA, 2004. MIT Press.

Phillip Bonacich. Factoring and weighting approaches to status scores and clique detection. *Journal of Mathematical Sociology*, pages 113–120, 1972.

Andrei Broder. Generating random spanning trees. In *FOCS '89: Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 442–447. Institute of Electrical and Electronics Engineers, 1989.

Alice Cheng and Eric Friedman. Manipulability of pagerank under sybil strategies. In *First Workshop on the Economics of Networked Systems (NetEcon06)*, 2006. URL `http://www.cs.duke.edu/nicl/netecon06/papers/ne06-sybil.pdf`.

Thomas H. Corman, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 2001.

N. Craswell and D. Hawking. Overview of the TREC 2003 Web track. In *The Twelfth Text Retrieval Conference*, number SP 500-255 in NIST Special Publication, 2003.

Madhav Desai, Sunil Kumar, and P. R. Kumar. Quasi-statically cooled markov chains. *Probability in the Engineering and Informational Sciences*, 8:1–19, 1994.

J. L. Doob. *Stochastic Processes*. John Wiley and Sons, Inc., New York, 1953.

Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 613–622, New York, NY, USA, 2001. ACM.

Nadav Eiron and Kevin S. McCurley. Links in hierarchical information networks. In L. Sandy Maisel, editor, *Lecture Notes in Computer Science (3243): Algorithms and Models for the Web-Graph*, pages 143–155. Springer–Verlag, 2004.

Mark Friedlin and Alexander Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer–Verlag, Berlin, 1984.

Anna Gambin and Piotr Pokarowski. A combinatorial aggregation algorithm for stationary distribution of a large markov chain. In *Fundamentals of Computation Theory*, volume 2138 of *Lecture Notes in Computer Science*, pages 384–387. Springer, 2001.

Eugene Garfield. Citation analysis as a tool in journal evaluation. *Science*, 178:471–479, 1972.

Zoltan Gyongyi and Hector Garcia-Molina. Web spam taxonomy. Technical report, Stanford University Technical Report, 2004.

T. Haveliwala and S. Kamvar. The second eigenvalue of the google matrix. Technical report, Stanford University Technical Report, 2003.

Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

Charles H. Hubbell. An input-output approach to clique identification. *Sociometry*, 28: 377–399, 1965.

Marius Iosifescu. *Finite Markov Processes and Their Applications*. John Wiley & Sons, New York, 1980.

S. Kamvar, T. Haveliwala, C. Manning, and G. Golub. Exploiting the block structure of the web for computing pagerank. Technical report, Stanford Digital Library Technologies Project, 2003a.

S. Kamvar, T. Haveliwala, C. Manning, and G. Golub. Exploiting the block structure of the web for computing pagerank. Technical report, Stanford University Technical Report, 2003b.

Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56, January 1993.

Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 1981.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43, 1953.

Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 95–103, New York, NY, USA, 2007. ACM.

Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 668–677, 1998.

Donald E. Knuth. *The Art of Computer Programming*, volume 1. Addison–Wesley, Reading, Massachusetts, third edition, 1997.

Jerrold E. Marsden. *Elementary Classical Analysis*. W. H. Freeman and Company, San Francisco, 1974.

Lawrence Page and Sergey Brin. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International World Wide Web Conference (WWW7)*, 1998.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

H. L. Royden. *Real Analysis*. Macmillan, New York, 1968.

Herbert A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962.

Frank W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer–Verlag, New York, 1984.

Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.

Richard L. Wheeden and Antoni Zygmund. *Measure and Integral: An Introduction to Real Analysis*. Marcel Dekker, Inc., New York, 1977.

John R. Wicks. *Linear algebra : an interactive laboratory approach with* Mathematica. Addison-Wesley, Reading, Mass., 1996.

H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, January 1993.