Abstract of "An Algorithm to Compute the Stochastically Stable Distribution of a Perturbed Markov Matrix" by John R. Wicks, Ph.D., Brown University, May, 2009.

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

May, 2009

This dissertation by John R. Wicks is accepted in its present form by
the Department of Computer Science as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date ＿＿＿＿＿＿＿   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿
                                     Amy Greenwald, Director

Recommended to the Graduate Council

Date ＿＿＿＿＿＿＿   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿
                                     Ugur Cetintemel, Reader

Date ＿＿＿＿＿＿＿   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿
                                     Roberto Serrano, Reader
                                         (Economics)

Approved by the Graduate Council

Date ＿＿＿＿＿＿＿   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿
                                   Dean of the Graduate School

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 transition matrix that satisfies the Markov property. The long-term behavior of such a Markov chain can be summarized by another probability distribution, which is a particular example of a *stable* distribution, or equilibrium. Under certain conditions, a Markov matrix has a unique stable distribution, which may be computed using standard linear algebra techniques. In general, however, a Markov matrix may have an infinite number of stable distributions, so that determining the long-term behavior of a Markov chain requires more difficult analysis.

Economists and game theorists also use finite-state, stationary Markov chains to study, for example, market dynamics and learning in repeated games. It is quite common for such models to have multiple equilibria. 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, and which has the added benefit of forcing any such model to converge to a unique long-term 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 to be 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* (or the *reduce* construction). 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 all but one member of a set of "closed" 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 reduction construction 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 reduction to develop novel proofs of classic results on the nature of the set of stable distributions of a Markov matrix.

We first present the reduce construction in the context of Markov *matrices*, deferring making the connection to 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. In chapters 4 and 5 we will show that we may 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. Likewise, chapters 2, 3, 4, and 5 will be largely dedicated to demonstrating the connection between M's closed sets and its collection of stable distributions.

Because the reduce 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, showing that a set of indices is open iff the corresponding submatrix of M is invertible. This then lays the groundwork for chapter 5 where we show how to eliminate open sets of indices, by applying our main construction, *reduction*, to reduce the dimension of M, without losing any information about its stable distributions. Also, in chapter 5, we present another important idea, which we call *scaling*. For unperturbed matrices, scaling may be recognized as right-preconditioning, a standard technique used to solve linear systems of equations.

In this chapter we also introduce 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 one is 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. However, we begin in chapter 6 by taking some time to prove the Markov Chain Tree Theorem (MCTT) in detail. Although it is not computationally practical as a means of

computing the SSD of a PMM, the MCTT provides the theoretical basis for most of chapter 7. To our knowledge, this proof is novel. Moreover, we feel it has a pleasing geometric flavor to it.

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 always be able to take limits as $\epsilon \to 0$ (i.e., continuity), and the unweighted graph of M_ϵ must be constant for $\epsilon > 0$. Thus, the entries of M_ϵ must be well-behaved, and they must remain so as such as we apply algebraic operations on M_ϵ. These 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.

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 defining *perturbed* matrices in section 7.2 (as matrices with exponentially convergent entries), we define perturbed *Markov* matrices PMMs in section 7.3 (as unichain Markov matrices with exponentially convergent entries). Using the MCTT, we show that the unique stable distribution of a PMM, M_ϵ, is a perturbed matrix, v_ϵ, so that its limit, v_0, as $\epsilon \to 0$, i.e., the stochastically stable distribution of M_ϵ, is well-defined. In sections 7.4 through 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 in the graph of $M_0 \equiv \lim_{\epsilon \to 0} M_\epsilon$, 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).

If we are instead interested in computing the *rate of convergence* to the SSD, we show, in chapter 8, how minor changes to the SSD algorithm allows us to calculate this in a similar manner. This problem has been studied in the context of Generalized Simulated Annealing, and the solution may be interpreted as an energy function. Thus, we call this the Energy algorithm. We prove that, unlike the SSD algorithm where we must work one communicating class at a time, we may reduce all communicating classes simultaneously. Although the algorithm of Gambin and Pokarowski (2001) only yields an approximation to the SSD, it is based on an exact calculation of the corresponding rates of convergence, which turns out to be the same as our Energy algorithm. However, their algorithm is mainly combinatorial, operating primarily as a recursive algorithm on graphs. Since our algorithm recursively operates on GSAs, we believe that it is conceptually more satisfying.

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 9, 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 the Energy algorithm of chapter 8 would output precisely the same solution as **GraphRank**. If true, this would give us a Markov chain interpretation for **GraphRank**, a la Dwork et al. (2001).

In chapter 10, we present another 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.

QuickRank is actually not simply a single algorithm, but a whole class of algorithms parameterized by a given *base* ranking algorithm, which we apply "locally," meaning 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 desirable axioms, the *peer-review principle* and *Bonacich's hypothesis*.

The idea of exploiting a hierarchy in this way was suggested 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, however, that BlockRank was developed as a first approximation to PageRank. In contrast, we argue that **QuickRank** may produce superior results, in that they more accurately reflect the judgements of local experts and are resistant to the ranking manipulation technique of web-spamming.

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. Usually, 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 ith element of σ by σ_i (instead of $\sigma(i)$).

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   Graph Theory Essentials

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, such as

Figure 1.1: Directed and Undirected Graphs
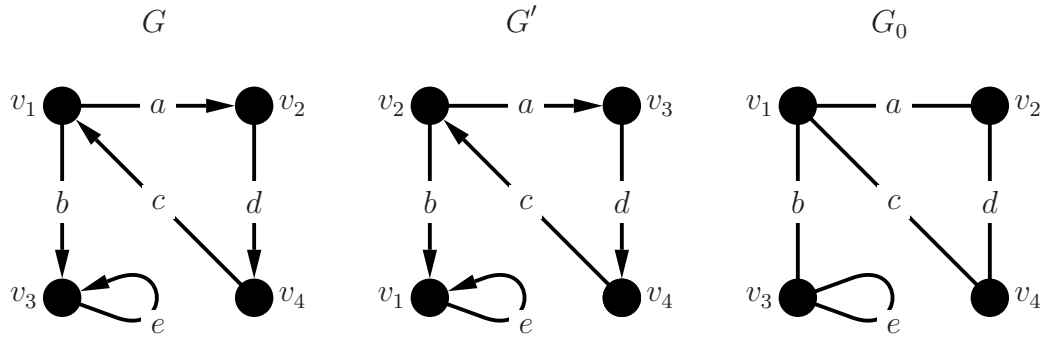
in in Figure 1.1. In this figure, vertices are drawn as geometric points, and edges as arrows from one vertex to another. Specifically, the arrow corresponding to an edge, $\alpha \in E$, starts at the point $s(\alpha)$ and terminates (i.e., ends) at $t(\alpha)$. If $v = s(\alpha)$ and $w = t(\alpha)$, we say that w is *adjacent* to s. For example, in G, $V = \{v_1, v_2, v_3, v_4\}$, $E = \{a, b, c, d, e\}$, $s(a) = v_1$, $t(a) = v_2$, etc., and v_2 is adjacent to v_1.

We will say two graphs, $G = (V, E, s, t)$ and $G' = (V', E', s', t')$, are *equal* iff there are 1-1 correspondences, $\delta : V' \to V$ and $\gamma : E' \to E$ such that $\delta s' = s\gamma$ and $\delta t' = t\gamma$, i.e., the connections in the graph are the same – only the labels on the elements of the graph are different. For example, the graphs G and G' in Figure 1.1 are equal with $\delta(v_1) = v_3$, $\delta(v_2) = v_1$, $\delta(v_3) = v_2$, $\delta(v_4) = v_4$, and $\gamma(\alpha) = \alpha$, $\forall \alpha \in E$.

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, such as G_0. Clearly, every graph has an associated undirected version, and, as in Figure 1.1, if G is a graph, we will denote its undirected version by G_0. If we define the *reverse*, G_R, of a directed graph, G, as we may define $G_0 = (G \cup G_R)_T$, where the union of two graphs is obtained by taking the unions of corresponding vertex and edge sets.

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. In general, for convenience, $\alpha = (v, w) \in E$ will denote an $\alpha \in E$ with $s(\alpha) = v$ and $t(\alpha) = w$, even when such an edge is not uniquely defined.

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. For example, the sequence, (c, a, d, c, b), specifies a walk in the graph, G, shown in Figure 1.1, from v_4 to v_3. By dropping edges, we obtain the corresponding path, (c, b), from v_4 to v_3.

In a graph that has no repeated edges, given an enumeration of its vertices, i.e., a 1-1 correspondence, $v : S_{|V|} \to V$, a walk may also be specified by a sequence $\sigma : S_l^0 \to S_{|V|}$ of $l + 1$ vertices. For example, the walk given above corresponds to the sequence, $\sigma = (4, 1, 2, 4, 1, 3)$, with associated path given by $\sigma' = (4, 1, 3)$. 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 $1 \leq i \leq l - 1$. As is clear from the examples, 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 essentially a path of length at least 1, except that we allow and require the initial and final vertices to be the same; that is, $\sigma_0 = \sigma_l$, or $s(\alpha_1) = t(\alpha_l)$. To convert the walk, σ, given above to its corresponding path, σ', we removed the cycle, $\sigma = (1, 2, 4, 1)$. A *self-loop* is a cycle of length 1, such as edge e in G.

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, i.e., each vertex has an associated self-loop. For example, in Figure 1.2, we show G and its transitive closure, G_T, although by convention we have suppressed drawing all the self-loops.

Figure 1.2: Unweighted and Weighted Graphs
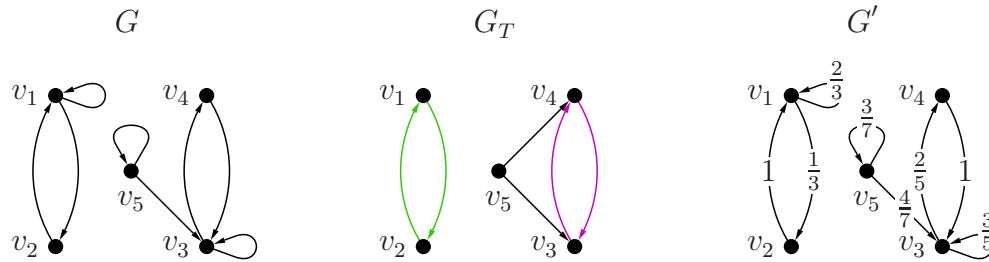
This yields a natural preorder[1] on V, given by the "leads to" relation, $\rightsquigarrow$, where $v_i \rightsquigarrow v_j$ iff $(v_i, v_j) \in E_T$. This preorder gives rise to an equivalence relation, $\sim$, where $i \sim j$ (read, "v_i is *strongly connected* to v_j") iff $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Equivalence classes with respect to $\sim$ are often called the *strongly connected components* (SCCs) of G.[2] For example, in Figure 1.2 the strongly connected components are given by $\{v_1, v_2\}$, $\{v_3, v_4\}$, and $\{v_5\}$, as indicated by the colored edges. Note that SCCs are maximal, meaning they do not contain other SCCs; furthermore, the SCCs of G *partition* the vertices of G, meaning each vertex belongs to exactly one SCC. A graph is said to be *strongly connected* if consists of only one strongly connected component. For example, a *complete* graph, which contains an edge from every vertex to every other vertex, is strongly connected.

Similarly, we have a "connects to" relation, $\leftrightsquigarrow$, associated with the undirected graph corresponding to G, G_0. That is, $v_i \leftrightsquigarrow v_j$ iff there is an *undirected* walk (i.e., a walk in the associated undirected graph) from v_i to v_j in G. The equivalence classes associated with this relation are called the *connected components* of G. For example, the connected components of G from Figure 1.2 are $\{v_1, v_2\}$, $\{v_3, v_4, v_5\}$. A graph is said to be *connected* if consists of only one connected component. A connected graph with no cycles is called a *tree*.

Notice that the connected components of G and G_T are exactly the same. Likewise, their strongly connected components are identical. Moreover, the (strongly) connected components in G_T are precisely its maximal complete subgraphs.

Thus far, we have restricted our attention to directed and undirected *unweighted* graphs.

---

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

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

Much of this thesis is actually concerned with weighted graphs. A weighted (directed or undirected) graph is one augmented with a function $d : E \rightarrow \mathbb{R}$, 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. An example of a weighted graph, G', is shown in Figure 1.2. Notice that the sum of the weights over the edges emanating from any given vertex is 1, so in this case, we may interpret the weights as the probability of traversing a given edge conditioned on the fact that we are at a particular vertex.

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, $G_- = (V, E)$. For example, with respect to the graphs in Figure 1.2, $G'_- = G$.

We will call a (directed/undirected, weighted/unweighted) graph G' a *subgraph* of a graph G, if the vertex and edge sets of G' are subset of those of G, with the corresponding restrictions of all other ancillary functions (e.g., d, s, t, etc.). For example, given a subset, V', of the vertex set of G, we define the *restriction* of G, to V', denoted $G|_{V'}$, to be the subgraph of G with vertex set, V', and the set of all edges with both ends in V'. If G' is a subgraph of G and their vertex sets are equal, we will say that G' is a *spanning* subgraph of G, or that G' *spans* G. A subgraph which is also a tree, it generally called a *subtree*. Thus, a *spanning subtree* refers to a subgraph which is a tree and which spans G.

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 $(v_i, v_j) \in E$, 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*. We say that a graph is *unichain* if it possesses exactly one chosed class.[3]

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

---

[3]This terminology comes from the theory of Markov chains (Iosifescu, 1980).

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 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 from v 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 through the vertices $s_i \in \mathcal{C}_{\sigma_i}$ and $t_i \in \mathcal{C}_{\sigma_{i+1}}$, with $\mathcal{C}_{\sigma_i} \neq \mathcal{C}_{\sigma_{i+1}}$.

Suppose we never encounter a closed class, i.e., $\mathcal{C}_{\sigma_i}$ is not closed for all i. Since there are only finitely many SCCs in G (i.e., $\sigma_i \in S_m$), for some $i < j$, we must have $\sigma_i = \sigma_j$. By construction, there is a walk from v to s_{j-1}, the starting vertex of the incoming edge to $\mathcal{C}_{\sigma_j}$. This is the concatenation of 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}. In particular, $s_i \rightsquigarrow s_{j-1}$. 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 $s_{j-1} \rightsquigarrow s_i$. This implies that $s_i \sim s_{j-1}$, which is a contradiction, since $\mathcal{C}_{\sigma_{j-1}} \neq \mathcal{C}_{\sigma_j} = \mathcal{C}_{\sigma_i}$. Thus, we must have $\mathcal{C}_{\sigma_i}$ closed for some i, and we have constructed a walk from v to $t_{i-1} \in \mathcal{C}_{\sigma_i}$. $\square$

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'. $\square$

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. Equivalently, the number of closed classes cannot increase as we add new edges. This observation will be particularly relevant in chapter 7.

Theorem 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 in $V' \setminus \mathcal{C}'$. 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. $\square$

Theorem 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 Theorem 1.3 with $G' = \overline{G}$ and $\mathcal{C}' = \overline{\mathcal{C}}$ to conclude that $\overline{\mathcal{C}}$ is a closed class of G. $\square$

Theorem 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 $\overline{G}|_{\mathcal{C}}$.[4] By Lemma 1.1, $\overline{G}|_{\mathcal{C}}$ 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 Theorem 1.3, with $G = \overline{G}$, $G' = \overline{G}|_{\mathcal{C}}$, and $\mathcal{C}' = \overline{\mathcal{C}}$, to conclude that $\overline{\mathcal{C}}$ is a closed class of $\overline{G}$ which is contained in $\mathcal{C}$. $\square$

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 ith row and j^{th} column of M. Observe that $M_{i,j} = e_i^t M e_j$, where $(e_i)_j = [i = j]$,[5] i.e., e_i has a 1 in the ith 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

---

[4]Note that, since $V = \overline{V}, \mathcal{C} \subset \overline{V}$, so that this restriction is defined.

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

is, for a column vector v, $v_i = v_{i,1}$, and for a row vector w, $w_j = w_{1,j}$. We will denote the set of column vectors of dimension $n \times 1$ by $\mathbb{R}^n$. 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 $0 \leq m < n$ rows and $0 \leq k < n$ columns of M to obtain an $(n - m) \times (n - k)$ 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,

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

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

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

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 will write $M \geq 0$ to indicate that its entries are non-negative, i.e., $M_{i,j} \geq 0, \ \forall i, j$. 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 $J = (1, \ldots, 1)$ 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[6] 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, $\Delta_n = \{x \geq 0 \mid \sum_{i=1}^n x_i = 1\}$, 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 $d(v_i, v_j) = M_{j,i}$. Notice that graphs obtained in this way cannot have repeated edges (cf. section 1.1). 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. When M is Markov, every vertex in $G(M)$ must have at least one outgoing edge. For example, the Markov matrix M on the left of Figure 1.3 gives rise to the "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 Theorem 1.5, if we increase the number of non-zero entries of M, the number of closed classes cannot increase

---

[6]Basic linear algebra concepts, such as this, that are not defined in the main body of the thesis are reviewed in Appendix A.

Figure 1.3: Markov Matrix and its associated "Markov" Weighted Graph

$$
M \qquad\qquad\qquad G\,(M)
$$

$$
\begin{pmatrix}
\frac{2}{3} & \frac{1}{4} & 0 & 0 & 0 \\
\frac{1}{3} & \frac{3}{4} & 0 & 0 & 0 \\
0 & 0 & \frac{3}{5} & \frac{1}{6} & \frac{4}{7} \\
0 & 0 & \frac{2}{5} & \frac{5}{6} & 0 \\
0 & 0 & 0 & 0 & \frac{3}{7}
\end{pmatrix}
$$
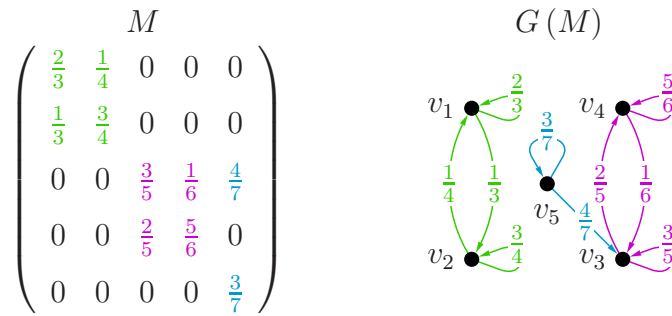
and must eventually decrease, since a complete graph consists of exactly one SCC, which is necessarily closed.

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 reduce construction we present in chapter 5 is defined in terms of sub-matrices, in this chapter, we carefully lay the groundwork for proving theorems about sub-matrices. 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 partitioning that isolates the submatrix, $M_{s,s}$. Further, we prove that $\imath_s$ is always injective, 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 a 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 $\overline{k} \equiv n - k$, 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 $\bar{s}_1 = 2, \bar{s}_2 = 3, \bar{s}^{-1}(2) = 1$, and $\bar{s}^{-1}(3) = 2$.

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 $\left(M_{s,s'} \right)_{i,j} = e_i^t \pi_s M \imath_{s'} e_j = (\imath_s e_i)^t M \imath_{s'} e_j = e_{s_i}^t M e_{s_j'} = M_{s_i, s_j'}$.

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. $\square$

2.2 Matrix Permutations

A *permutation* of a set s is a bijective mapping $s \rightarrow s$. 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, we may define the associated *permutation matrix*, P^ρ, so that $P^\rho_{i,j} = [i = \rho(j)]$. Notice that a permutation matrix is invertible with inverse equal to its transpose.

Theorem 2.2. *If ρ is a permutation of S_n, then $(P^\rho)^{-1} = (P^\rho)^t = P^{(\rho^{-1})}$.*

Proof. Applying the definition above, gives

$$(P^\rho)^t_{i,j} = (P^\rho)_{j,i} = [j = \rho(i)] = [\rho^{-1}(j) = i] = P^{(\rho^{-1})}_{i,j}$$

Moreover,

$$\left(P^{(\rho^{-1})} P^\rho\right)_{i,j} = \sum_{k=1}^n \left(P^{(\rho^{-1})}\right)_{i,k} (P^\rho)_{k,j}$$

$$= \sum_{k=1}^n [i = \rho^{-1}(k)] \, [k = \rho(j)] = [i = \rho^{-1}(\rho(j))] = [i = j] = I_{i,j}$$

Therefore, $P^{(\rho^{-1})} P^\rho = I$. Since P^ρ is a square matrix, by Theorem A.2, $P^{(\rho^{-1})} = (P^\rho)^{-1}$. $\square$

Likewise, the *matrix permutation* of an $n \times n$ matrix, M, according to ρ is defined as $M' = (P^\rho)^t M P^\rho$. This is simply a rearrangement of the entries of M.

Theorem 2.3. *Given a permutation, ρ, of S_n and an $n \times n$ matrix M, if $M' = (P^\rho)^t M P^\rho$, then $M'_{i,j} = M_{\rho(i),\rho(j)}$.*

Proof. Using Theorem 2.2,

$$\left(M'\right)_{i,j} = \left((P^\rho)^t M P^\rho\right)_{i,j} = \sum_{k,l} \left((P^\rho)^t\right)_{i,k} M_{k,l} P^\rho_{l,j}$$

$$= \sum_{k,l} (P^\rho)_{k,i} M_{k,l} [l = \rho(j)] = \sum_k ((P^\rho))_{k,i} M_{k,\rho(j)}$$

$$= \sum_k [k = \rho(i)] M_{k,\rho(j)} = M_{\rho(i),\rho(j)}$$

$\square$

Figure 2.1: The Effect of a Matrix Permutation on a Markov Graph

$$M$$

$$\begin{pmatrix} 0 & 0 & 0 & 0.4 \\ 0.1 & 0 & 0 & 0 \\ 0.2 & 0 & 0.3 & 0 \\ 0 & 0.5 & 0 & 0 \end{pmatrix}$$

$$M'$$

$$\begin{pmatrix} 0.3 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.4 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$$

$$G(M)$$

$$G(M')$$
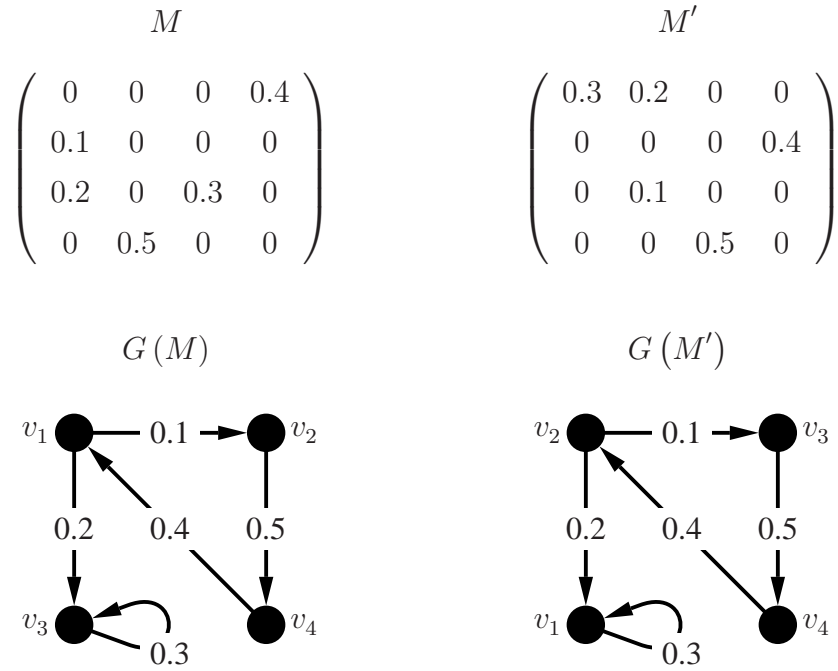
For example, consider the permutation, ρ, given by $\rho(1) = 3$, $\rho(2) = 1$, $\rho(3) = 2$, and $\rho(4) = 4$. Applying the corresponding matrix permutation to M given in Figure 2.1 yields the matrix M', also in the Table. Notice that the corresponding graphs, $G(M)$ and $G(M')$, are identical, up to a relabelling of the vertices.

Using the submatrix construction given in section 2.1, for any subset $s \subset S_n$, we can define a permutation matrix, P_s, 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.

Theorem 2.4. *Given $s \subset S_n$, if $\rho_s : S_n \to S_n$ is given by*

$$\rho_s(i) = \begin{cases} \overline{s}_i & \text{if } i \leq \overline{k} \\ s_{i-\overline{k}} & \text{otherwise} \end{cases}$$

then $P_s = \begin{pmatrix} \imath_{\overline{s}} & \imath_s \end{pmatrix}$ is a permutation matrix with $P_s = P^{\rho_s}$.

Proof. First observe that ρ_s is clearly surjective, since $S_n = s \cup \overline{s}$. Since S_n is a finite set, this must be 1-1 as well. In particular, ρ_s is a permutation of S_n. To show that $P_s = P^{\rho_s}$,

we proceed by cases. If $j \leq \overline{k}$, then

$$
\begin{aligned}
(P_s)_{i,j} &= e_i^t P_s e_j = e_i^t \left(\begin{array}{cc} \imath_{\overline{s}} & \imath_s \end{array} \right) e_j = e_i^t \imath_{\overline{s}} e_j = e_i^t e_{\overline{s}_j} \\
&= \left[i = \overline{s}_j \right] = \left[i = \rho_s(j) \right] = P_{i,j}^{\rho_s}
\end{aligned}
$$

Similarly, if $j > \overline{k}$,

$$
(P_s)_{i,j} = e_i^t \left(\begin{array}{cc} \imath_{\overline{s}} & \imath_s \end{array} \right) e_j = e_i^t e_{s_{j-\overline{k}}} = \left[i = s_{j-\overline{k}} \right] = \left[i = \rho_s(j) \right] = P_{i,j}^{\rho_s}
$$

Since their entries are equal, $P_s = P^{\rho_s}$. $\square$

Now notice that if $P_s = \left(\begin{array}{cc} \imath_{\overline{s}} & \imath_s \end{array} \right)$, then $P_s^t = \left(\begin{array}{c} \pi_{\overline{s}} \\ \pi_s \end{array} \right)$, and

$$
P_s^t M P_s = \left(\begin{array}{c} \pi_{\overline{s}} \\ \pi_s \end{array} \right) M \left(\begin{array}{cc} \imath_{\overline{s}} & \imath_s \end{array} \right) = \left(\begin{array}{cc} \pi_{\overline{s}} M \imath_{\overline{s}} & \pi_{\overline{s}} M \imath_s \\ \pi_s M \imath_{\overline{s}} & \pi_s M \imath_s \end{array} \right) = \left(\begin{array}{cc} M_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & M_{s,s} \end{array} \right).
$$

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.

Example 2.5. With s and M as in Example 2.1,

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

so that M has the partitioning $M_{s,s} = \left(\begin{array}{cc} 1 & 4 \\ 13 & 16 \end{array} \right)$, $M_{\overline{s},s} = \left(\begin{array}{cc} 5 & 8 \\ 9 & 12 \end{array} \right)$, $M_{s,\overline{s}} = \left(\begin{array}{cc} 2 & 3 \\ 14 & 15 \end{array} \right)$,

and $M_{\overline{s},\overline{s}} = \left(\begin{array}{cc} 6 & 7 \\ 10 & 11 \end{array} \right)$ with respect to $s = \{1, 4\}$. $\square$

Corollary 2.6. *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, $J P_s = J$. Therefore, $J =$

$JP_sP_s^{-1} = JP_s^{-1}$. However, by Theorems 2.2 and 2.4, $P_s^{-1} = P_s^t$, so that $J = JP_s^{-1} = JP_s^t$, i.e., 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 $P_s^tMP_s$, when M is $n \times n$ and Markov. $\square$

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 left-multiplication by $\imath_s$ is a mapping $\mathbb{R}^k \to \mathbb{R}^n$. We call $\imath$ an *inclusion* operator, because it is a 1-1 linear transformation which maps $\mathbb{R}^k$ to a subspace of $\mathbb{R}^n$. In fact, for $v \in \mathbb{R}^k$, $w = \imath_sv \in \mathbb{R}^n$ 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.

Similarly, π_s yields a mapping $\mathbb{R}^n \to \mathbb{R}^k$. This is a surjective linear transformation and hence corresponds to an orthogonal *projection*. Specifically, for $v' \in \mathbb{R}^n$, the coordinates of $w' = \pi_sv' \in \mathbb{R}^k$ are just the coordinates of v' with indices in s.

Example 2.7. Let $s = \{1,4\} \subset S_4$ and $v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4$. The vector $w = \pi_sv =$

$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ is the corresponding projection of v. The vector $u =$

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

Next, we will define $\mathbb{R}^s \equiv \text{span}\{e_i \mid i \in s\}$, a subspace of $\mathbb{R}^n$ of dimension k. Simi-larly, $\mathbb{R}^{\overline{s}} \equiv \text{span}\{e_i \mid i \in \overline{s}\}$ is a subspace of $\mathbb{R}^n$ of dimension $\overline{k}$. The next lemma high-lights the key algebraic and geometric properties of these projection and inclusion opera-tors.

Lemma 2.8. *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) $\ker \imath_{\overline{s}} = 0 = \ker \imath_s$, $\text{im } \pi_{\overline{s}} = \mathbb{R}^{\overline{k}}$, *and* $\text{im } \pi_s = \mathbb{R}^k$;

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

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 injec-tive with $\ker \imath_{\overline{s}} = \ker \imath_s = \{0\}$. Likewise, $\pi_{\overline{s}}$ and π_s are right-invertible, hence surjective with $\text{im } \pi_{\overline{s}} = \mathbb{R}^{\overline{k}}$ and $\text{im } \pi_s = \mathbb{R}^k$.

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

$$\text{im } \imath_s = \left\{ \imath_s v \mid v \in \mathbb{R}^k \right\} \qquad \text{by definition of im}$$

$$= \left\{ \imath_s \sum_{i=1}^k v_i e_i \mid v_i \in \mathbb{R}, e_i \in \mathbb{R}^k \right\} \qquad \text{since } \{e_i\} \text{ is a basis}$$

$$= \left\{ \sum_{i=1}^k v_i \imath_s e_i \mid v_i \in \mathbb{R}, e_i \in \mathbb{R}^k \right\} \qquad \text{by linearity of matrix multiplication}$$

$$= \left\{ \sum_{i=1}^k v_i e_{s_i} \mid v_i \in \mathbb{R}, e_{s_i} \in \mathbb{R}^n \right\} \qquad \text{by by defintion of } \imath_s$$

$$= \text{span}\left\{ e_j \mid j \in s \right\} \qquad \text{by definition of span}$$

$$= \mathbb{R}^s \qquad \text{by definition of } \mathbb{R}^s$$

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

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

□

Notice that 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 $v = v_s + v_{\bar{s}}$, with $v_s \in \mathbb{R}^s$ and $v_{\bar{s}} \in \mathbb{R}^{\bar{s}}$.

Example 2.9. Take s and v as in Example 2.7. Define

$$v_s = \imath_s\pi_s v = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 4 \end{pmatrix}.$$

Similarly, define

$$v_{\bar{s}} = \imath_{\bar{s}}\pi_{\bar{s}} v = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 3 \\ 0 \end{pmatrix}.$$

Now observe that $v_s \in \mathbb{R}^s$, $v_{\bar{s}} \in \mathbb{R}^{\bar{s}}$, and $v_s + v_{\bar{s}} = v$. □

Theorem 2.10. *Given $s \subset S_n$, $\mathbb{R}^n = \mathbb{R}^{\bar{s}} \oplus \mathbb{R}^s$. Further, for any $v \in \mathbb{R}^n$, $v = v_{\bar{s}} + v_s$, with $v_{\bar{s}} = \imath_{\bar{s}}\pi_{\bar{s}}v \in \mathbb{R}^{\bar{s}}$ and $v_s = \imath_s\pi_s v \in \mathbb{R}^s$.*

Proof. Let $v \in \mathbb{R}^{\bar{s}} \cap \mathbb{R}^s$, so that $v = \sum_{i \in \bar{s}} a_i e_i$, and $v = \sum_{j \in s} b_j e_j$, and $\sum_{i \in \bar{s}} a_i e_i - \sum_{j \in s} b_j e_j = 0$. Since the standard basis vectors are linearly independent, $a_i = 0 = b_j$ for all i and j, and $v = 0$. In other words, $\mathbb{R}^{\bar{s}} \cap \mathbb{R}^s = 0$.

Now for any $v \in \mathbb{R}^n$, by Lemma 2.8 a), $v = Iv = (\imath_{\bar{s}}\pi_{\bar{s}} + \imath_s\pi_s) v = \imath_{\bar{s}}\pi_{\bar{s}}v + \imath_s\pi_s v = v_{\bar{s}} + v_s$. Since $\operatorname{im} \imath_{\bar{s}} = \mathbb{R}^{\bar{s}}$ and $\operatorname{im} \imath_s = \mathbb{R}^s$, it follows that $v_{\bar{s}} \in \mathbb{R}^{\bar{s}}$ and $v_s \in \mathbb{R}^s$. But then $v \in \mathbb{R}^{\bar{s}} + \mathbb{R}^s$, so that $\mathbb{R}^n = \mathbb{R}^{\bar{s}} + \mathbb{R}^s$. Therefore, $\mathbb{R}^n = \mathbb{R}^{\bar{s}} \oplus \mathbb{R}^s$. □

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 2.11. *Let M be an $n \times n$ Markov matrix, and an invariant set of indices, $s \subset S_n$.*

   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 invariant. So there are no edges in $G(M)$ from s to $\overline{s}$. This 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. Furthermore, $M_{s,s} \geq 0$, since $M, \imath_s, \pi_s \geq 0$. 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 since $\operatorname{im} \imath_s = \mathbb{R}^s$ and $\operatorname{im} \imath_{s'} = \mathbb{R}^{s'}$, so $\operatorname{im} \imath_s \subset \operatorname{im} \imath_{s'}$. 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:

$$M_{s',s'}v' = \pi_{s'}M\imath_{s'}v' = \pi_{s'}M\imath_s v = \pi_{s'}IM\imath_s v$$

$$= \pi_{s'}\left(\imath_{\overline{s}}\pi_{\overline{s}} + \imath_s\pi_s\right)M\imath_s v \qquad \text{by Lemma 2.8 a)}$$

$$= \pi_{s'}\imath_{\overline{s}}\pi_{\overline{s}}M\imath_s v + \pi_{s'}\imath_s\pi_s M\imath_s v$$

$$= \pi_{s'}\imath_{\overline{s}}M_{\overline{s},s}v + \pi_{s'}\imath_s M_{s,s}v$$

$$= 0 + \pi_{s'}\imath_s M_{s,s}v \qquad \text{by part a)}$$

$$= \pi_{s'}\imath_s v \qquad \text{by assumption}$$

$$= \pi_{s'}\imath_{s'}v'$$

$$= v' \qquad \text{by Lemma 2.8 a)}$$

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 $\imath_{s'} = I$. Thus, $v' = \imath_s v \in I \ker \Lambda = \ker \Lambda$. Hence, we need only show that v' is a distribution. Since $v, \imath_s \geq 0$, it follows that $v' \geq 0$. Moreover, $Jv' = J\imath_s v = J\sum_{i=1}^{|s|} v_i e_{s_i} = \sum_{i=1}^{|s|} v_i J e_{s_i} = \sum_{i=1}^{|s|} v_i = \sum_{i=1}^{|s|} v_i J e_i = J\sum_{i=1}^{|s|} v_i e_i = Jv = 1$, so that $v' \in \operatorname{stab} M$.

$\square$

Theorem 2.12. *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 $\dim \ker \Lambda^t \neq 0$. 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 2.11 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 2.11 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 2.8 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 2.11 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 2.11 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 $i \rightarrow \infty$, 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 $N \rightarrow \infty$. More generally, for any Markov matrix, M^j converges in the Cesaro sense (Doob, 1953; Iosifescu, 1980), that is, the matrix, $M^\infty \equiv \lim_{N \rightarrow \infty} \frac{1}{N}\sum_{j=0}^{N-1} M^j$ is well-defined. Since $v_i = M^i v_0$, this implies that v_i converges in the Cesaro sense to $\lim_{N \rightarrow \infty} \frac{1}{N}\sum_{i=0}^{N-1} v_i = \lim_{N \rightarrow \infty} \frac{1}{N}\sum_{i=0}^{N-1} M^i v_0 = M^\infty v_0$.

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, π_{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 $\text{stab}(M)$ is the set of convex combinations of the columsn of M^∞. In particular, this proves the existence of a stable distribution for any Markov matrix. Moreover, this will allow us to give an explicit formula for M^∞, in section 5.2, that does not involve limits.

Lemma 3.1. *If M is Markov, then*

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

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

    c) *there are unique, well-defined linear mappings, $\pi_{\text{ker}} : \mathbb{R}^n \rightarrow \ker \Lambda$ and $\pi_{\text{im}} : \mathbb{R}^n \rightarrow \text{im}\,\Lambda$ such that $\pi_{\text{ker}} + \pi_{\text{im}} = I$.*

Proof. We first show that $\ker \Lambda \cap \operatorname{im} \Lambda = 0$. Take $v \in \ker \Lambda \cap \operatorname{im} \Lambda$. Since $v \in \operatorname{im} \Lambda$, 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 enough 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 $W = \operatorname{im} \Lambda$, and using the fact that $V \cap W = 0$, we have that $\dim(\ker \Lambda + \operatorname{im} \Lambda) = \dim \ker \Lambda + \dim \operatorname{im} \Lambda$. By Theorem A.1, $\dim \operatorname{im} M + \dim \ker M = n$ for any $m \times n$-dimensional matrix, M. Thus, $\dim(\ker \Lambda + \operatorname{im} \Lambda) = n$. Since $\ker \Lambda + \operatorname{im} \Lambda \subset \mathbb{R}^n$, we must have $\ker \Lambda + \operatorname{im} \Lambda = \mathbb{R}^n$, so that $\ker \Lambda \oplus \operatorname{im} \Lambda = \mathbb{R}^n$.

This means that the mapping $\alpha : \ker \Lambda \times \operatorname{im} \Lambda \to \mathbb{R}^n$, such that $\alpha(v, w) = v + w$, for $v \in \ker \Lambda$ and $w \in \operatorname{im} \Lambda$, 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 $w \in \operatorname{im} \Lambda$, so $v, w \in \ker \Lambda \cap \operatorname{im} \Lambda = 0$. 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, $\alpha^{-1} : \mathbb{R}^n \to \ker \Lambda \times \operatorname{im} \Lambda$, corresponding to a pair of linear mappings, $\pi_{\ker}$ and π_{im}, with $\operatorname{im} \pi_{\ker} \subset \ker \Lambda$ and $\operatorname{im} \pi_{\operatorname{im}} \subset \operatorname{im} \Lambda$. In particular, $\alpha(\pi_{\ker}, \pi_{\operatorname{im}}) = \pi_{\ker} + \pi_{\operatorname{im}} = I$, since $v = \left(\alpha\alpha^{-1}\right)v = \alpha\left(\alpha^{-1}v\right) = \alpha\left(\pi_{\ker}v, \pi_{\operatorname{im}}v\right) = \pi_{\ker}v + \pi_{\operatorname{im}}v = \left(\pi_{\ker} + \pi_{\operatorname{im}}\right)v$. Since inverses are unique, $\pi_{\ker}$ and π_{im} are the unique such mappings. $\square$

We now give formulas for $\pi_{\ker}$ and π_{im} directly in terms of M. In particular, we will see that the Cesaro limit of Markov matrices mentioned at the beginning of this section exists. In addition, we find that $\pi_{\ker}$ is Markov.

Theorem 3.2. *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 $\pi_{\ker}$ as $N \to \infty$. In other words, particular, M^∞ exists and equals $\pi_{\ker}$. In particular, $\pi_{\ker}$ is Markov and $I - M^\infty = \pi_{\operatorname{im}}$.*

Proof. We begin by showing 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). Observe that, since $0 \leq \left\| M^j \right\|_1 \leq 1$, $\left\| \sum_{j=0}^{N-1} M^j \right\|_1 \leq \sum_{j=0}^{N-1} \left\| M^j \right\|_1 \leq N$, so that $\| M_N \|_1 \leq 1$. In particular, M_N is a bounded sequence, so that 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, $v = \pi_{\text{ker}} v + \pi_{\text{im}} v$. Fix a basis of $\ker \Lambda$, i.e., a spanning set of independent, stable vectors with respect to M (and hence, with respect to M^i for all i), $\{v_1, \ldots, v_k\}$. We may then write $\pi_{\text{ker}} v = \sum_{r=1}^k \beta_r v_r \in \ker \Lambda$ for some $\beta_r \in \mathbb{R}^k$, and $\pi_{\text{im}} v = \Lambda w$ for some $w \in \mathbb{R}^n$. Applying, M_{N_i} then gives:

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

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}, and we may write $M^\infty = \pi_{\text{ker}}$. Since $I = \pi_{\text{ker}} + \pi_{\text{im}}$, we automatically have that $\pi_{\text{im}} = I - \pi_{\text{ker}} = I - M^\infty$.

Finally, we argue that M^∞, and hence π_{ker}, is Markov. Observe that, for any N, M_N is Markov. It is non-negative by definition, and $J M_N = J \frac{1}{N} \sum_{j=0}^{N-1} M^j = \frac{1}{N} \sum_{j=0}^{N-1} J M^j = \frac{1}{N} \sum_{j=0}^{N-1} J = J$. Since $M_N \geq 0$, $M^\infty = \lim_{N \to \infty} M_N \geq \lim_{N \to \infty} 0 = 0$. Moreover, $J M^\infty = J \lim_{N \to \infty} M_N = \lim_{N \to \infty} J M_N = \lim_{N \to \infty} J = J$. $\square$

We close this chapter with our first structure theorem in which we characterize the stable distributions of M in terms of M^∞. Specifically, we prove the classical result (Iosifescu, 1980, Theorem 5.3, p. 155) that each column of M^∞ is a stable distribution of M, and every stable distribution is a convex combination of the columns of M^∞. In particular, every Markov matrix, M, has at least one stable distribution.

Corollary 3.3. *For any Markov matrix,* M, $\mathrm{stab}(M) = \mathrm{stab}\,(M^\infty) = M^\infty \Delta_n$. *In partic-ular, for all* i, $M^\infty e_i \in M^\infty \Delta_n = \mathrm{stab}\,(M^\infty) = \mathrm{stab}(M)$.

Proof. First, observe that if $v \in \ker \Lambda$, since $\pi_{\mathrm{ker}} v + \pi_{\mathrm{im}} v = v$, $\pi_{\mathrm{im}} v = v - \pi_{\mathrm{ker}} \in \ker \Lambda \cap \mathrm{im}\,\Lambda$, so that $0 = \pi_{\mathrm{im}} v$ and $v = \pi_{\mathrm{ker}} v$. By Theorem 3.2, $M^\infty = \pi_{\mathrm{ker}}$ and $I - M^\infty = \pi_{\mathrm{im}}$. By our convention, we will write Λ^∞ for $M^\infty - I = -\pi_{\mathrm{im}}$, so that $\ker \Lambda^\infty = \ker\,(I - M^\infty) = \ker \pi_{\mathrm{im}}$.

Now observer that, if $v \in \ker \Lambda^\infty$, $v = M^\infty v = \pi_{\mathrm{ker}} v \in \ker \Lambda$. Conversely, if $v \in \ker \Lambda$, $0 = \pi_{\mathrm{im}} v = -\Lambda^\infty v$, and $v \in \ker \Lambda^\infty$. In other words, $\ker \Lambda^\infty = \ker \Lambda$, from which it follows that $\mathrm{stab}\,(M^\infty) = \ker \Lambda^\infty \cap \Delta_n = \ker \Lambda \cap \Delta_n = \mathrm{stab}(M)$.

In addition, by Theorem 3.2, M^∞ is Markov, so that $M^\infty \Delta_n \subset \Delta_n$. Since $M^\infty \Delta_n \subset \mathrm{im}\,M^\infty = \mathrm{im}\,\pi_{\mathrm{ker}} \subset \ker \Lambda$, it follows that $M^\infty \Delta_n \subset \ker \Lambda \cap \Delta_n = \mathrm{stab}(M)$. Conversely, assume that $v \in \mathrm{stab}(M)$, so that $v \in \ker \Lambda$. By our earlier observation, $M^\infty v = \pi_{\mathrm{ker}} v = v$. In particular, since $v \in \Delta_n$, $v \in M^\infty \Delta_n$. Thus, $\mathrm{stab}(M) = M^\infty \Delta_n$. $\square$

Example 3.4. 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 3.3 predicts. $\square$

Chapter 4

Sub-Markov Matrix Invertibility

We begin this chapter by defining various 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 $\mathcal{S}_n(i,j,l) = \left\{\sigma \in \mathcal{S}_n(l) \mid \sigma_0 = j \text{ and } \sigma_l = i\right\}$ 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 $\mathcal{S}_n(s,i,j,l) = \left\{\sigma \in \mathcal{S}_n(i,j,l) \mid \sigma_t \in s, \, \forall\, 0 < t < l\right\}$ to be the set of sequences in S_n of length $l + 1$ from j to i whose intermediate values all lie in s. However, 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 $V_s \equiv \left\{v_i \in V \mid i \in s\right\}$ 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. Now $\bigcup_{i \in S_n} \mathcal{P}_M(i, j, n)$ corresponds to walks in $G(M)$ of length n starting at v_j. Likewise, $\bigcup_{i \in s} \mathcal{P}_M(s, i, j, n)$ corresponds to those same walks with the additional condition that they end in s and include only vertices in s. Thus, $\bigcup_{i \in s} \mathcal{P}_M(s, i, j, n)$ is a subset of $\bigcup_{i \in S_n} \mathcal{P}_M(i, j, n)$

Since 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 $\bigcup_{i \in S_n} \mathcal{P}_M(i, j, n)$. $\square$

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 $\sigma \in \mathcal{P}_M(l)$.*

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., $\sigma \in \mathcal{P}_M(l)$.

$\square$

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. We may similarly identify sums over various subsets of walks in terms of products of submatrices of M.

Lemma 4.3. *For any $n \times n$ matrix M and index sets $\alpha, \beta, s \subset S_n$,*

a) $\left(M^l\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M(i,j,l)} W(M,\sigma)$, *and*

b) $\left(M_{\alpha,s} M_{s,s}^{l-2} M_{s,\beta}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M(s,\alpha_i,\beta_j,l)} W(M,\sigma)$,

Proof.

Proof of part a): Using the facts that $M_{i,j} = e_i^t M e_j$ and $I = \sum_{k=1}^{n} e_k e_k^t$, 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,

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

where the final equality follows from Lemma 4.2.

Proof of part b): Similarly, if $k = |s|$,

$$
\begin{aligned}
\left(M_{\alpha,s} M_{s,s}^{l-2} M_{s,\beta} \right)_{i,j} &= e_i^t M_{\alpha,s} \left(\sum_{r_{l-1}=1}^{k} e_{r_{l-1}} e_{r_{l-1}}^t \right) M_{s,s} \dots M_{s,s} \left(\sum_{r_1=1}^{k} e_{r_1} e_{r_1}^t \right) M_{s,\beta} e_j \\
&= \sum_{r_{l-1}=1}^{k} \dots \sum_{r_1=1}^{k} e_i^t M_{\alpha,s} e_{r_{l-1}} e_{r_{l-1}}^t M_{s,s} \dots M_{s,s} e_{r_1} e_{r_1}^t M_{s,\beta} e_j \\
&= \sum_{r_{l-1}=1}^{k} \dots \sum_{r_1=1}^{k} M_{\alpha_i, s_{r_{l-1}}} M_{s_{r_{l-1}}, s_{r_{l-2}}} \dots M_{s_{r_2}, s_{r_1}} M_{s_{r_1}, \beta_j}.
\end{aligned}
$$

This time each choice of values for the summation variables, $\{r_1, \dots, r_{l-1}\}$, represents a unique choice of $\sigma \in \mathcal{S}_n(s, \alpha_i, \beta_j, l)$, and the conclusion follows as in part a).

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

$\square$

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.

Theorem 4.4. *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 $i \geq 0$, $\left\| M_{s,s}^i \right\|_1 \leq c^{\lfloor \frac{i}{n} \rfloor}$ for some $0 \leq c < 1$, and $\lim_{i \to \infty} M_{s,s}^i = 0$.*

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

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 $c^{\lfloor \frac{i}{n} \rfloor}$. Remember from section 1.2 that for a positive matrix, $M \geq 0$, $\|M\|_1$ is just its maximum column sum. Since M is Markov, and since $M_{s,s}$ is a submatrix of M, $0 \leq \left\|M_{s,s}\right\|_1 \leq \|M\|_1 \leq 1$. 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 all column sums of $M_{s,s}^n$ are strictly less than 1. Since $JM = J$, $JM^n = J$, i.e., all column sums of M^n equal 1. In particular, writing out the s_j^{th} column sum of M^n, we have:

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

Equation 4.2 follows from Lemma 4.3 a). The inequality in 4.3 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. Next, we re-index and apply Lemma 4.3 b) with $\alpha = \beta = s$ to arrive at Equation 4.4, which is the jth column sum of $M_{s,s}^n$, showing that this is strictly less than 1. Since j was arbitrary, $\left\|M_{s,s}^n\right\|_1 < 1$.

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), $\sum_{i=0}^{\infty} \left\| M_{s,s}^i \right\|_1 \leq \sum_{i=0}^{\infty} c^{\lfloor \frac{i}{n} \rfloor}$. By writing $i = kn + j$ for $0 \leq j < n$, we may rewrite this as $\sum_{k=0}^{\infty} \sum_{j=0}^{n-1} c^{\lfloor \frac{kn+j}{n} \rfloor} = \sum_{k=0}^{\infty} \sum_{j=0}^{n-1} c^k = n \sum_{k=0}^{\infty} c^k = \frac{n}{1-c}$. Hence, the summation $\sum_{i=0}^{\infty} \left\| M_{s,s}^i \right\|_1$ is bounded above and is, hence convergent. Since each entry of $\sum_{i=0}^{N} M_{s,s}^i$ is an increasing sequence which is bounded above by $\sum_{i=0}^{\infty} \left\| M_{s,s}^i \right\|_1$, they each converge, i.e., 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 $I - M_{s,s}$. Since both matrices are square, by Theorem A.2, it suffices to show that the sum is a right-inverse, as follows:

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

where the final step in the derivation follows from part a).

Proof of part c): Inversely, assume s contains an entire closed class of M. By Theorem 2.12 b), $\ker \Lambda_{s,s} \neq 0$. But $I - M_{s,s} = -(M_{s,s} - I) = -\Lambda_{s,s}$, 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 2.12 b), $\ker \Lambda_{s,s} \neq 0$. 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, $\lim_{i \to \infty} M_{s,s}^i \neq 0$.

$\square$

Corollary 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) $\left(-M_{\bar{s},s} \Lambda_{s,s}^{-1}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,\bar{s}_i,s_j\right)} W(M, \sigma),$

b) $\left(-\Lambda_{s,s}^{-1} M_{s,\bar{s}}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,s_i,\bar{s}_j\right)} W(M, \sigma),$ *and*

c) $\left(M_{\overline{s},\overline{s}} - M_{\overline{s},s} \Lambda_{s,s}^{-1} M_{s,\overline{s}}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j\right)} W(M,\sigma).$

Proof.

Proof of part a): By Theorem 4.4 b), $-M_{\overline{s},s} \Lambda_{s,s}^{-1} = M_{\overline{s},s}\left(I - M_{s,s}\right)^{-1} = \sum_{i=0}^{\infty} M_{\overline{s},s} M_{s,s}^i$. Since projecting to the $(i,j)^{\text{th}}$ index is linear, applying Lemma 4.3 b) with $\alpha = \overline{s}$ and $\beta = s$ gives

$$
\begin{aligned}
\left(-M_{\overline{s},s}\Lambda_{s,s}^{-1}\right)_{i,j} &= \left(\sum_{k=0}^{\infty} M_{\overline{s},s}M_{s,s}^k\right)_{i,j} = \sum_{k=0}^{\infty}\left(M_{\overline{s},s}M_{s,s}^k\right)_{i,j} \\
&= \sum_{k=0}^{\infty}\sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_i,s_j,k+1)} W(M,\sigma) = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,s_j\right)} W(M,\sigma)
\end{aligned}
$$

Proof of part b): Similarly, $-\Lambda_{s,s}^{-1} M_{s,\overline{s}} = \sum_{i=0}^{\infty} M_{s,s}^i M_{s,\overline{s}}$ so that applying Lemma 4.3 b) with $\alpha = s$ and $\beta = \overline{s}$ gives

$$
\left(-\Lambda_{s,s}^{-1}M_{s,\overline{s}}\right)_{i,j} = \sum_{k=0}^{\infty}\sum_{\sigma \in \mathcal{P}_M(s,s_i,\overline{s}_j,k+1)} W(M,\sigma) = \sum_{\sigma \in \mathcal{P}_M\left(s,s_i,\overline{s}_j\right)} W(M,\sigma).
$$

Proof of part c): Finally,

$$
\left(M_{\overline{s},\overline{s}} - M_{\overline{s},s}\Lambda_{s,s}^{-1}M_{s,\overline{s}}\right)_{i,j} = \left(M_{\overline{s},\overline{s}}\right)_{i,j} + \sum_{k=0}^{\infty}\left(M_{\overline{s},s}M_{s,s}^k M_{s,\overline{s}}\right)_{i,j}.
$$

Notice that $\mathcal{P}_M\left(\overline{s}_i,\overline{s}_j,1\right) = \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j,1\right)$ is a singleton set, such that $\left(M_{\overline{s},\overline{s}}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j,1\right)} W(M,\sigma)$. By Lemma 4.3 b) with $\alpha = \beta = \overline{s}$, $\left(M_{\overline{s},s}M_{s,s}^k M_{s,\overline{s}}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j,k+2\right)} W(M,\sigma)$, so that

$$
\left(M_{\overline{s},\overline{s}} - M_{\overline{s},s}\Lambda_{s,s}^{-1}M_{s,\overline{s}}\right)_{i,j} = \sum_{l=1}^{\infty}\sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j,l\right)} W(M,\sigma) = \sum_{\sigma \in \mathcal{P}_M\left(s,\overline{s}_i,\overline{s}_j\right)} W(M,\sigma).
$$

$\square$

Chapter 5

Two Useful Constructions

In this chapter, we present our two fundamental constructions, scale and reduce. 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 reduce construction will allow us to sharpen the structure theorem shown in Corollary 3.3 by proving Theorem 5.17 and Theorem 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 $\mathrm{stab}\,(M_1)$. $\square$

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 $D_{i,i}^{-1} = (D_{i,i})^{-1}$). In particular, it has both left and right inverses, and is thus both injective and surjective.

For example,

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

These correspond to the graphs in Figure 5.1.

Figure 5.1: $G(M)$ and $G(M_D)$ with $D = \text{diag}\left(\frac{1}{3}, \frac{3}{2}, \frac{1}{2}\right)$.
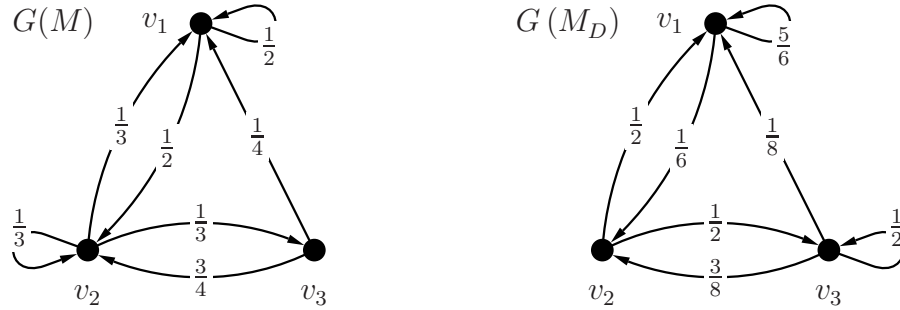
By looking at the corresponding graphs, we see that scaling by D adjusts the weight of the self-loop at each vertex with a proportional adjustment of the weights on the corresponding outgoing edges. When $D_{i,i}$ is close to 0, the weight of the resulting self-loop at i is large (near 1), and when $D_{i,i}$ is close to $\left(1 - M_{i,i}\right)^{-1}$, the the weight of the resulting self-loop is small (near 0).

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. By Lemma A.3 b), therefore, $\ker \Lambda = D \ker \Lambda D = D \ker \Lambda_D$. That is, $M_D \approx_D M$. $\square$

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 $\epsilon \max_i (1 - M_{i,i}) \leq 1$, 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 $M_\epsilon \approx M$. $\square$

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:

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

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

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

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 now compute 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 $\Lambda_{s,s} = \begin{pmatrix} -\frac{1}{2} \end{pmatrix}$, $\Lambda_{s,s}^{-1} = \begin{pmatrix} -2 \end{pmatrix}$, $M_{s,\bar{s}} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}$,

$$M_{\bar{s},s} = \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix}. \text{ So we can calculate } \imath = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \; p = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}, \; \widehat{M} =$$

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

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

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

Proof.

Proof of part a): By Theorem 4.4 4.4,

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

$M_{s,\bar{s}}$ and $M_{\bar{s},s}$ 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, $JM_{s,s} + JM_{\bar{s},s} = J$. Therefore, $JM_{\bar{s},s} = J - JM_{s,s} = JI - JM_{s,s} = -J\Lambda_{s,s}$, so that

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

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 $\mathbb{R}^k$. Thus, $\dim \operatorname{im} p = k$ and p is surjective. 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, by Theorem A.1, $\operatorname{rk} \imath + \dim \ker \imath = k$, so $\dim \ker \imath = 0$, and $\imath$ is injective.

Proof of part c): This follows from Equations 5.1 -5.3 and matrix algebra.

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

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

---

[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.

Theorem 5.8. *If M is Markov and $s \in S_n$ an open set of indices in M with $|s| = k$,*

a) $e_i^t p = e_{\bar{s}_i}^t + \sum_{j=1}^{k} \sum_{\sigma \in \mathcal{P}_M\left(s, \bar{s}_i, s_j\right)} W(M, \sigma) e_{s_j}^t,$

b) $\imath e_j = e_{\bar{s}_j} + \sum_{i=1}^{k} \sum_{\sigma \in \mathcal{P}_M\left(s, s_i, \bar{s}_j\right)} W(M, \sigma) e_{s_i},$

c) $\left(\widehat{M}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s, \bar{s}_i, \bar{s}_j\right)} W(M, \sigma),$ *and*

d) $\mathcal{P}\left(\widehat{M}\right) = \mathcal{P}\left(M\right)|_{V_{\bar{s}}},$ *where v_i in $\mathcal{P}\left(\widehat{M}\right)$ corresponds to $v_{\bar{s}_i}$ in $\mathcal{P}\left(M\right)$.*

Proof.

Proof of part a): First, observe that from Equation 5.2,

$$
p = \left(I \quad -M_{\bar{s},s}\, \Lambda_{s,s}^{-1} \right) P_s^t = \left(I \quad -M_{\bar{s},s}\, \Lambda_{s,s}^{-1} \right) \begin{pmatrix} \pi_{\bar{s}} \\ \pi_s \end{pmatrix}
$$

$$
= \pi_{\bar{s}} - M_{\bar{s},s}\, \Lambda_{s,s}^{-1} \pi_s
$$

As in the proof of Lemma 4.3, we may use the fact that $I = \sum_{j=1}^{k} e_j e_j^t$ to compute the ith row of p as

$$
\begin{aligned}
e_i^t p &= e_i^t \pi_{\bar{s}} - e_i^t M_{\bar{s},s}\, \Lambda_{s,s}^{-1} \pi_s = e_{\bar{s}_i}^t - e_i^t M_{\bar{s},s}\, \Lambda_{s,s}^{-1} \sum_{j=1}^{k} e_j e_j^t \pi_s \\
&= e_{\bar{s}_i}^t + \sum_{j=1}^{k} e_i^t \left(-M_{\bar{s},s}\, \Lambda_{s,s}^{-1}\right) e_j e_{s_j}^t = e_{\bar{s}_i}^t + \sum_{j=1}^{k} \left(-M_{\bar{s},s}\, \Lambda_{s,s}^{-1}\right)_{i,j} e_{s_j}^t \\
&= e_{\bar{s}_i}^t + \sum_{j=1}^{k} \sum_{\sigma \in \mathcal{P}_M\left(s, \bar{s}_i, s_j\right)} W(M, \sigma) e_{s_j}^t
\end{aligned}
$$

where we have appealed to Corollary 4.5 a) for the final equality.

Proof of part b): The proof of this part is similar to that of part a). We first observe that

$$
\imath = P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\bar{s}} \end{pmatrix} = \left(\imath_{\bar{s}} \quad \imath_s \right) \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1} M_{s,\bar{s}} \end{pmatrix} = \imath_{\bar{s}} - \imath_s \Lambda_{s,s}^{-1} M_{s,\bar{s}}.
$$

Therefore, we may compute the jth column of $\imath$ as

$$
\begin{aligned}
\imath e_j &= \imath_{\bar{s}} e_j - \imath_s \Lambda_{s,s}^{-1} M_{s,\bar{s}} e_j = e_{\bar{s}_j} - \imath_s \sum_{i=1}^{k} e_i e_i^t \Lambda_{s,s}^{-1} M_{s,\bar{s}} e_j \\
&= e_{\bar{s}_j} + \sum_{i=1}^{k} \imath_s e_i e_i^t \left(-\Lambda_{s,s}^{-1} M_{s,\bar{s}} \right) e_j = e_{\bar{s}_j} + \sum_{i=1}^{k} \left(-\Lambda_{s,s}^{-1} M_{s,\bar{s}} \right)_{i,j} e_{s_i} \\
&= e_{\bar{s}_j} + \sum_{i=1}^{k} \sum_{\sigma \in \mathcal{P}_M \left(s, s_i, \bar{s}_j \right)} W(M, \sigma) e_{s_i}
\end{aligned}
$$

appealing to Corollary 4.5 b) at the end.

Proof of part c): Since $\widehat{M} = M_{\bar{s},\bar{s}} - M_{\bar{s},s} \Lambda_{s,s}^{-1} M_{s,\bar{s}}$, this part follows immediately from Corollary 4.5 c).

Proof of part d): Finally, 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 $i, j \in \bar{s}$, this walk may then be decomposed into a concatenation of walks whose interior vertices are in V_s which originate and terminate in $V_{\bar{s}}$. 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, $\mathcal{P}\left(\widehat{M}\right) = \mathcal{P}(M)|_{V_{\bar{s}}}$.

$\square$

Example 5.9. 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 $s = \{3\}$ to be $\widehat{M} = \begin{pmatrix} 0 & \frac{3}{4} \\ 1 & \frac{1}{4} \end{pmatrix}$.

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. Specifically, $\widehat{M}_{1,2}$ is the sum of the weight of the edge (v_2, v_1) and the weights of all paths which begin at v_2, cycle at v_3 i times, and end at v_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, $\widehat{M}_{2,2} = \sum_{i=0}^{\infty} M_{2,3}(M_{3,3}^i)M_{3,2} = \sum_{i=0}^{\infty} \frac{1}{4}(\frac{1}{2})^i \frac{1}{2} = \frac{1}{4}$. $\square$

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

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

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 were to have more than one closed class, at least one of them would have to be contained in s, which is impossible since s is open. 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 $\tilde{s} \equiv S_n - \{i\} - s$, then we may apply Theorem 5.10 to conclude that $\overline{s}^{-1}(\tilde{s})$ is open with respect to $\widehat{M}$. Since $\tilde{s} \subset \overline{s}$, $|\tilde{s} \cap \overline{s}| = |\tilde{s}| = |\overline{s}| - 1$, 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 $\mathcal{P}(\widehat{M}) = \mathcal{P}(M)|_{V_{\overline{s}}}$, $\mathcal{P}(\widehat{M})$ must be complete, as well. In particular, $\widehat{M}$ is irreducible. $\square$

We will show in section 5.3 that if we consider a Markov chain, 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 reduce 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: $A_1 = \begin{pmatrix} I & M_{\overline{s},s} \\ 0 & I \end{pmatrix}$, $A_2 = \begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s} \end{pmatrix}$, and $A_3 = P_s$. Now A_3 has inverse P_s^t, and it is easy to check that A_1 has inverse $\begin{pmatrix} I & -M_{\overline{s},s} \\ 0 & I \end{pmatrix}$. In addition, since s is open, Theorem 4.4 implies that $\Lambda_{s,s} = -(I - M_{s,s})$ is invertible, so that A_2 has inverse

$\begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s}^{-1} \end{pmatrix}$. Therefore,

$$\begin{aligned}
A_3 A_2 A_1 B &= P_s \begin{pmatrix} I & 0 \\ 0 & \Lambda_{s,s} \end{pmatrix} \begin{pmatrix} I & M_{\bar{s},s} \\ 0 & I \end{pmatrix} \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} \\
&= P_s \begin{pmatrix} \widehat{\Lambda} & M_{\bar{s},s} \\ 0 & \Lambda_{s,s} \end{pmatrix} \\
&= P_s \begin{pmatrix} \Lambda_{\bar{s},\bar{s}} - M_{\bar{s},s}\Lambda_{s,s}^{-1}M_{s,\bar{s}} & M_{\bar{s},s} \\ 0 & \Lambda_{s,s} \end{pmatrix} \qquad (5.4) \\
&= P_s \begin{pmatrix} \Lambda_{\bar{s},\bar{s}} & M_{\bar{s},s} \\ M_{s,\bar{s}} & \Lambda_{s,s} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\bar{s}} & I \end{pmatrix} \\
&= P_s \left(P_s^t \Lambda P_s \right) \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\bar{s}} & I \end{pmatrix} \qquad (5.5) \\
&= \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\bar{s}} & I \end{pmatrix}.
\end{aligned}$$

Equation 5.4 follows from Theorem 5.6 c), and Equation 5.5 follows from the definitions of P_s and the component submatrices of Λ. Now since A_1, A_2, A_3 are invertible, in particular they are injective. So by Lemma A.3 c), $\ker A_3 A_2 A_1 B = \ker B$. That is,

$\ker \Lambda P_s \begin{pmatrix} I & 0 \\ -\Lambda_{s,s}^{-1}M_{s,\bar{s}} & I \end{pmatrix} = \ker \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix}$.

Now consider Lemma 2.8 with $s' = S_n - S_{|\bar{s}|}$, so that $\imath_{\bar{s}'} = \begin{pmatrix} I \\ 0 \end{pmatrix}$, $\imath_{s'} = \begin{pmatrix} 0 \\ I \end{pmatrix}$,

$\pi_{\bar{s}'} = \begin{pmatrix} I & 0 \end{pmatrix}$, and $\pi_{s'} = \begin{pmatrix} 0 & I \end{pmatrix}$. Since $\begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix} \widehat{\Lambda} \begin{pmatrix} I & 0 \end{pmatrix} +$

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

$$\ker \begin{pmatrix} \widehat{\Lambda} & 0 \\ 0 & I \end{pmatrix} = \ker \left[\begin{pmatrix} I \\ 0 \end{pmatrix} \widehat{\Lambda} \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} 0 & I \end{pmatrix} \right]$$

$$= \ker \left[\imath_{\overline{s}'} \widehat{\Lambda} \pi_{\overline{s}'} + \imath_{s'} \pi_{s'} \right]$$

$$= \ker \left[\imath_{\overline{s}'} \widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \ker \left[\imath_{s'} \pi_{s'} \right] \qquad \text{by Lemma A.3 d)}$$

$$= \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \ker \pi_{s'} \qquad \text{by Lemma A.3 c)}$$

$$= \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \right] \cap \operatorname{im} \imath_{\overline{s}'} \qquad \text{by Lemma 2.8 c)}$$

$$= \imath_{\overline{s}'} \ker \left[\widehat{\Lambda} \pi_{\overline{s}'} \imath_{\overline{s}'} \right] \qquad \text{by Lemma A.3 a)}$$

$$= \imath_{\overline{s}'} \ker \widehat{\Lambda} . \qquad \text{by Lemma 2.8 a)}$$

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. $\square$

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 $\mathrm{stab}\left(\widehat{M}\right)$ to $\mathrm{stab}(M)$.*

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, $\dim \ker \Lambda = k$. $\square$

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 3.3, $|\mathrm{stab}\, M| > 1$. More specifically, by Lemma 2.11 d), $\mathrm{stab}\, M$ contains $\imath_s \overline{v}$, where $\overline{v} \in \mathrm{stab}\, M_{s,s}$, the stable distribution of the principal submatrix, $M_{s,s}$. Since M is unichain, by Theorem 5.14, $\dim \ker \Lambda = 1$. Thus, if $v, w \in \mathrm{stab}\, M \subset \ker \Lambda$, we would have $v = kw$. However, since $1 = Jv = kJw = k$, $v = w$, so that $|\mathrm{stab}\, M| = 1$, namely, $\mathrm{stab}\, M = \{\imath_s \overline{v}\}$.

Letting $v = \imath_s \overline{v}$, consider $s' = \{i \mid v_i \neq 0\}$. It remains to show that $v_i \neq 0 \iff i \in s$, or equvalently, that $s = s'$. Consider any $i \notin s$. By Lemma 2.8 c), $v \in \mathrm{im}\, \imath_s = \mathbb{R}^s = \mathrm{span}\{e_i \mid i \in s\}$. Since $i \notin s$, $v_i = 0$, and $i \notin s'$. Thus, $\overline{s} \subset \overline{s}'$, so that $s' \subset s$.

Now suppose $i \in s$, but $i \notin s'$. Since $v \in \operatorname{stab} M$ and $v_j = 0$ for all $j \notin s'$, $0 = v_i = (Mv)_i = \sum_{j \in S_n} M_{i,j} v_j = \sum_{j \in s'} M_{i,j} v_j$. Since $v_j > 0$ for all $j \in s'$, and $M_{i,j} \geq 0$ for all $j \in s'$, we must have $M_{i,j} = 0$ for all $j \in s'$. That is, there are no edges from any $j \in s'$ to i in $G(M)$. Since the choice of $i \in s - s'$ was arbitrary, there are no edges from any $j \in s'$ to any $i \in s - s'$. But $s' \subset s$, and s is a closed class, so each j is in the same SCC as each i, and this is a contradiction. Therefore $s' = s$, or in other words, $v_i \neq 0 \iff i \in s$. $\square$

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, $v > 0$.*

While the next structure theorem for Markov matrices follows from the proof of Theorem 2.1 in Karlin and Taylor (1981, p. 4), the reduce construction provides a conceptually satisfying, constructive proof.

Theorem 5.17. *Given a Markov matrix M, with closed classes, $\{s^r\}_{r=1}^k$, let M_{s^r,s^r} be the principal submatrix on s^r with unique stable distribution, $\overline{v}^r > 0$. Let $v^r = \imath_{s^r} \overline{v}^r$, the inclusion of $\overline{v}^r$ in $\mathbb{R}^n$.*

a) $\ker \Lambda = \operatorname{span} \left\{ v^1, \ldots, v^k \right\}$,

b) $\ker \Lambda = \operatorname{span} \{ \operatorname{stab}(M) \}$, *and*

c) *every $w \in \operatorname{stab}(M)$ is a convex combination of the vectors v^r, i.e., $w = \sum_{j=1}^k \alpha_j v^j$ for $0 \leq \alpha_j \leq 1$ with $\sum_{j=1}^k \alpha_j = 1$.*

Proof. We should first observe that, for any t, by Lemma 2.11, M_{s^r,s^r} is Markov. Moreover, since $G_- \left(M_{s^r,s^r} \right)$ is strongly connected, M_{s^r,s^r} is irreducible. Therefore, by Corollary 5.16, M_{s^r,s^r} has a unique stable distribution, $\overline{v}^r > 0$.

Proof of part a): 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

$$
LD = \begin{pmatrix} J\pi_{s^1}v^1 & \cdots & J\pi_{s^1}v^k \\ \vdots & \ddots & \vdots \\ J\pi_{s^k}v^1 & \cdots & J\pi_{s^k}v^k \end{pmatrix}
$$

$$
= \begin{pmatrix} J\pi_{s^1}\imath_{s^1}\overline{v}^1 & \cdots & J\pi_{s^1}\imath_{s^k}\overline{v}^k \\ \vdots & \ddots & \vdots \\ J\pi_{s^k}\imath_{s^1}\overline{v}^1 & \cdots & J\pi_{s^k}\imath_{s^k}\overline{v}^k \end{pmatrix}
$$

$$
= \begin{pmatrix} J\overline{v}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J\overline{v}^k \end{pmatrix}
$$

$$
= I,
$$

because by Lemma 2.8, for two sets $s, s' \subset S_n$, $\pi_s \imath_{s'} = I$ when $s = s'$ and $\pi_s \imath_{s'} = 0$ when $s \cap s' = \emptyset$. So D is left-invertible with non-negative left inverse L.

In particular, D is injective. D is a $n \times k$ matrix with $\ker D = 0$, so $\dim \operatorname{im} D = k$. Now $\operatorname{im} D = \operatorname{span}\left\{v^1, \ldots, v^k\right\} \subset \ker \Lambda$ because by Lemma 2.11 d), for any column v^r of D, $v^r = \imath_{s^r}\overline{v}^r \in \operatorname{stab}(M) \subset \ker \Lambda$. But by Theorem 5.14, $\dim \ker \Lambda = k$. So $\ker \Lambda = \operatorname{im} D = \operatorname{span}\left\{v^1, \ldots, v^k\right\}$.

Proof of part b): Since $\left\{v^1, \ldots, v^k\right\} \subset \operatorname{stab}(M) \subset \ker \Lambda$,

$$
\operatorname{span}\left\{v^1, \ldots, v^k\right\} \subset \operatorname{span}\{\operatorname{stab}(M)\} \subset \ker \Lambda.
$$

Therefore, $\operatorname{span}\left\{v^1, \ldots, v^k\right\} = \operatorname{span}\{\operatorname{stab}(M)\} = \ker \Lambda$.

Proof of part c): If we view the $k \times k$ identity matrix, I, as a Markov matrix, then $\Lambda(I)$ is the zero matrix, and $\ker \Lambda(I) = \mathbb{R}^k$. This means that $\operatorname{stab} I = \ker \Lambda(I) \cap \Delta_k = \mathbb{R}^k \cap \Delta_k = \Delta_k$. Now $D \ker \Lambda(I) = \operatorname{im} D = \ker \Lambda$, and D is non-negative with non-negative left-inverse, so $I \approx_D M$, and by Lemma 5.2, D^* is a bijection from $\operatorname{stab}(I)$ to $\operatorname{stab}(M)$. But $JD = \begin{pmatrix} Jv^1 & \cdots & Jv^k \end{pmatrix} = J$, so that for any $v \in \Delta_k$, $\|Dv\|_1 = JDv = Jv = 1$, and $D^* = D$ on Δ_k. That is, D gives a bijection between Δ_k and $\operatorname{stab} M$. In particular, every element w in $\operatorname{stab} M$ satisfies $w = Dv$ for some $v \in \Delta_k$, so w is a convex combination of the vectors v^r.

□

We may now give a further characterization of the reduction when for each closed class either s does not interset it at all or s contains almost all of it. Notice this includes the case when s is maximal.

Theorem 5.18. *Using the notation of Theorem 5.17, assume that $s \subset S_n$ has the property that if $s^r \cap s \neq \emptyset$, $|s^r \cap \overline{s}| = 1$, and in this case define $\alpha'(r)$ and $\alpha(r)$ so that the unique entry of this singleton is given by $s^r_{\alpha'(r)} = \overline{s}_{\alpha(r)}$. If $(\widehat{M}, p, \imath)$ is the reduction of M with respect to s, if $j = \alpha(r)$, $\widehat{M}e_j = e_j$ and $\imath e_j = (v^r)^{-1}_{\overline{s}_j} v^r$.*

Proof. Assume that i is such that $\mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right) \neq \emptyset$ for $\overline{s}_j$. By assumption, $j = \alpha(r)$ and $\overline{s}_j \in s^r$. Since s^r is closed, we must have $\overline{s}_i \in s^r$, as well. However, since $\overline{s} \cap s^r$ is a singleton set, we must have $i = j = \alpha(r)$. Consequently, if $i \neq j$, $\mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right) = \emptyset$. Therefore, by Theorem 5.8 c), if $|s| = n - l$,

$$\widehat{M}e_j = \sum_{i=1}^{l} e_i e_i^t \widehat{M}e_j = \sum_{i=1}^{l} e_i \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right)} W(M, \sigma) = e_j \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_j, \overline{s}_j\right)} W(M, \sigma)$$

that is, the jth column of $\widehat{M}$ has only one non-zero entry in the jth row. Since $\widehat{M}$ is Markov, this entry must be 1, so that $\widehat{M}e_j = e_j$. This means that $e_j \in \text{stab}\left(\widehat{M}\right)$. By Theorem 5.12, $\imath$ maps stable distributions of $\widehat{M}$ to stable vectors of M. Thus, by Theorem 5.17, $\imath e_j = \sum_{r'=1}^{k} \beta_{r'} v^{r'}$ for some β_r.

Now observe that if $i \notin s^r$, $e_i^t \imath e_j = 0$. We can see this by considering the two cases, $s_i \notin s^r$ and $\overline{s}_i \notin s^r$. First, consider the case when $s_i \notin s^r$. As before, we know that $\mathcal{P}_M\left(s, s_i, \overline{s}_j\right) = \emptyset$, since $\overline{s}_j \in s^r$ and s^r is closed. Therefore, by Theorem 5.8 b), $e_{s_i}^t \imath e_j = \sum_{\sigma \in \mathcal{P}_M\left(s, s_i, \overline{s}_j\right)} W(M, \sigma) = 0$. In the second case, when $\overline{s}_i \notin s^r$, $e_{\overline{s}_i}^t \imath e_j = [i = j]$. Since $\overline{s}_j \in s^r$, $i \neq j$ and $e_{\overline{s}_i}^t \imath e_j = 0$.

If $i \notin s^{r'}$, $e_i^t \imath_{s^{r'}} = 0$, so that $e_i^t v^{r'} = e_i^t \imath_{s^{r'}} \overline{v}^{r'} = 0$. Therefore, if $i = s_{i'}^{r'}$,

$$e_i^t \imath e_j = \sum_{r''=1}^{k} \beta_{r''} e_i^t v^{r''} = \beta_{r'} e_i^t v^{r'} = \beta_{r'} \left(v^{r'}\right)_i.$$

If, in addition, $r' \neq r$, we have also just shown that $e_i^t \imath e_j = 0$ and since, by Corollary 5.15, $\left(v^{r'}\right)_i = e_i^t v^{r'} = e_i^t \imath_{s^{r'}} \overline{v}^{r'} = e_i^t \sum_{\overline{i}} \imath_{s^{r'}} e_{\overline{i}} e_{\overline{i}}^t \overline{v}^{r'} = e_i^t \sum_{\overline{i}} e_{s_{\overline{i}}^{r'}} e_{\overline{i}}^t \overline{v}^{r'} = e_{i'}^t \overline{v}^{r'} = \left(\overline{v}^{r'}\right)_{i'} > 0$, we must have $\beta_{r'} = 0$. In other words, $\imath e_j = \beta_r v^r$.

If $i = s^r_{\alpha'(r)} = \overline{s}_j$, Theorem 5.8 b) implies that $e^t_i \imath e_j = e^t_{\overline{s}_j} \imath e_j = 1$. Therefore, $1 = e^t_i \imath e_j = \beta_r e^t_i v^r = \beta_r (v^r)_i$, so that $\imath e_j = (v^r)^{-1}_i v^r = (v^r)^{-1}_{\overline{s}_j} v^r$. $\square$

In the special case when s contains no transient indices, we can give a simple formula for $\widehat{M}$.

Theorem 5.19. *Using the assumptions and notation of Theorem 5.18, assume further that $s \subset \bigcup_{r=1}^k s^r$. Then p simply sums the entries in each closed class intersecting s, i.e.,*

$$p_{i,j} = \begin{cases} [j = \overline{s}_i] & i \notin \operatorname{im}\alpha \\ \\ [j \in s^r] & \text{if } i = \alpha(r) \end{cases} \tag{5.6}$$

$$\tag{5.7}$$

Similarly, $\widehat{M}$ is obtained by dropping the columns and summing the rows in s, so that

$$\left(\widehat{M}\right)_{i,j} = \begin{cases} M_{\overline{s}_i,\overline{s}_j} & \text{if } i,j \neq \operatorname{im}\alpha \\ \\ \sum_{l \in s^r} M_{l,\overline{s}_j} & i = \alpha(r) \text{ and } j \neq \operatorname{im}\alpha \\ \\ [i = j] & j = \alpha(r) \end{cases} \tag{5.8}$$

Proof. By Theorem 5.8 a),

$$p_{i,j} = e^t_i p e_j = [j = \overline{s}_i] + \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_i,j)} W(M,\sigma)\,[j \in s]. \tag{5.9}$$

If $i \in \operatorname{im}\alpha$ with $i = \alpha(r)$ and $j \notin s^r$, $j \neq \overline{s}_i = s^r_{\alpha'(r)}$. If $j \notin s$, then both terms of Equation 5.9 vanish. If $j \notin s^r$, since $s^{r'}$ is closed, $\mathcal{P}_M(s,\overline{s}_i,j) = \emptyset$, and again Equation 5.9 equals 0.

Similarly, in the case when $i \neq \operatorname{im}\alpha$ and $j \neq \overline{s}_i$, we can show that Equation 5.9 vanishes. When $j \notin s$, this is trivially true. If $j \in s$, assume that $\mathcal{P}_M(s,\overline{s}_i,j) \neq \emptyset$. As before, we must have $j,\overline{s}_i \in s^r$ for some r. By assumption, since $j \in s^r \cap s$ and $\overline{s}_i \in s^r \cap \overline{s}$, we must have $i = \alpha(r)$, which is contrary to assumption. Therefore, $\mathcal{P}_M(s,\overline{s}_i,j) = \emptyset$, and again Equation 5.9 equals 0.

To this point, we have shown that the only non-vanishing entries in p occur in those entries (i,j) for which either $i = \alpha(r)$ and $j \in s^r$ or $i \neq \operatorname{im}\alpha$ and $j = \overline{s}_i$. Now observe that this implies that p has exactly one non-zero row in each column. Assume the contrary,

that there exists $i \neq i'$ and j, such that the $(i, j)^{\text{th}}$ and $(i', j)^{\text{th}}$ entries of p are both non-zero. There are three cases to consider. If $i, i' \in \operatorname{im} \alpha$, then $i = \alpha(r)$ and $i' = \alpha(r)$ for $r \neq r'$. For both the $(i, j)^{\text{th}}$ and $(i', j)^{\text{th}}$ entries of p to be non-zero, we must have $j \in s^r$ and $j \in s^{r'}$, but since distinct closed classes are disjoint, this is impossible. Another possibility is that neither i nor i' are in $\operatorname{im} \alpha$, in which case, we must have $j = \overline{s}_i$ and $j = \overline{s}_{i'}$, which is impossible. Finally, we must consider the case when $i = \alpha(r)$ and $i' \notin \operatorname{im} \alpha$, in which case $j \in s^r$ and $j = \overline{s}_{i'}$, so that $\overline{s}_{i'} \in s^r$. Since $\overline{s}_{\alpha(r)} = s^r_{\alpha'(r)}$, we also have $\overline{s}_i \in s^r$. However, for this choice of r, $|s^r \cap \overline{s}| = 1$, which would imply that $\overline{s}_i = \overline{s}_{i'}$, so that $i = i'$, a contradiction. Thus, p has exactly one non-zero row in each column. Since by Theorem a), the columns of p sum to 1, that non-zero entry must be 1, so that Equation 5.6 holds. Notice this implies that $\sum_{\sigma \in \mathcal{P}_M(s, \overline{s}_i, j)} W(M, \sigma) = 1$ for $i = \alpha(r)$ and $j \in s^r$.

To prove Equation 5.8, first observe that the case when $j = \overline{s}_i$ has already been proven in Theorem 5.18. Now consider the case when $i, j \neq \operatorname{im} \alpha$, and observe that $\mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right) \subset \mathcal{S}_n(1)$. If $\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right)$ were a path of length greater than 1, it would pass through s, in particular, it would pass through some closed class, s^r. This would force $\overline{s}_i \in s^r$, as well. But since $s^r \cap s \neq \emptyset$, $|s^r \cap \overline{s}| = 1$ so that $i = \alpha(r)$, contrary to assumption. Thus, $\mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right)$ consists of the unique path of length 1 starting at $\overline{s}_j$ and ending at $\overline{s}_i$, for which $W(M, \sigma) = M_{\overline{s}_i, \overline{s}_j}$. Thus, by Theorem 5.8 c),

$$\left(\widehat{M}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right)} W(M, \sigma) = \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j, 1\right)} W(M, \sigma) = M_{\overline{s}_i, \overline{s}_j}$$

Finally, consider the case when $i = \alpha(r)$ and $j \neq \operatorname{im} \alpha$, so that in particular $i \neq j$. Partitioning the set of paths into those of length 1 versus those of greater length which pass through a specific index in s gives

$$\left(\widehat{M}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right)} W(M, \sigma)$$

$$= \sum_{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j, 1\right)} W(M, \sigma) + \sum_{l=1}^{|s|} \sum_{\substack{\sigma \in \mathcal{P}_M\left(s, \overline{s}_i, \overline{s}_j\right) \\ \sigma(1) = s_l}} W(M, \sigma)$$

$$= M_{\overline{s}_i, \overline{s}_j} + \sum_{l=1}^{|s|} \sum_{\sigma \in \mathcal{P}_M(s, \overline{s}_i, s_l)} W(M, \sigma) M_{s_l, \overline{s}_j}$$

Since s_l is in some closed class, $\mathcal{P}_M\left(s, \overline{s}_i, s_l\right) \neq \emptyset$ only if it is in the same class as $\overline{s}_i$,

namely s^r. If $s^r \cap s = \emptyset$, then $\left(\widehat{M}\right)_{i,j} = M_{\bar{s}_i, \bar{s}_j}$. Otherwise, $s^r \cap s = s^r \setminus \{\bar{s}_i\}$, in which case

$$
\begin{aligned}
\left(\widehat{M}\right)_{i,j} &= M_{\bar{s}_i, \bar{s}_j} + \sum_{\substack{l \in s^r \\ l \neq \bar{s}_i}} \sum_{\sigma \in \mathcal{P}_M(s, \bar{s}_i, l)} W(M, \sigma) M_{l, \bar{s}_j} \\
&= M_{\bar{s}_i, \bar{s}_j} + \sum_{\substack{l \in s^r \\ l \neq \bar{s}_i}} M_{l, \bar{s}_j} \\
&= \sum_{l \in s^r} M_{l, \bar{s}_j}
\end{aligned}
$$

where we have used the earlier observation that $\sum_{\sigma \in \mathcal{P}_M(s, \bar{s}_i, j)} W(M, \sigma) = 1$ for $i = \alpha(r)$ and $j \in s^r$ $\square$

Example 5.20. We illustrate Theorem 5.19 using the same matrix as in Example 3.4, $M = \begin{pmatrix} 0 & 1 & 0 & \frac{1}{6} \\ 1 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$. This matrix has two closed classes, $\{1, 2\}$ and $\{3\}$. Reducing M with

respect to the open set, $s = \{2\}$ yields a 3-dimensional matrix with $\imath = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & \frac{1}{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and

$p = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Notice that the first two columns of $\imath$ do correspond to the stable

distributions of the two closed classes. Likewise, p sums entries 1 and 2 in any distribution,

as Theorem 5.19 says it should. Finally,

$$
\widehat{M} \;=\; p\Lambda\imath + I = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & \frac{1}{6} \\ 1 & -1 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix} \imath + I
$$

$$
= \begin{pmatrix} 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix} \imath + I = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & \frac{1}{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + I
$$

$$
= \begin{pmatrix} 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{6} \\ 0 & 0 & -\frac{1}{2} \end{pmatrix} + I = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{1}{6} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}
$$

The graph of M and its reduction are shown in Figure 5.2. This example illustrates that, under the assumptions of Theorem 5.19, $G\left(\widehat{M}\right)$ is just the graph of $G(M)$ with each set of vertices, V_{s^r}, "collapsed" to its representative, $v_{\alpha(t)}$, in $G\left(\widehat{M}\right)$ when $s^r \cap s \neq \emptyset$. $\square$

Theorem 5.21. *Using notation of Theorem 5.18, consider the case when s is maximal, so that $j = \operatorname{im} \alpha$ for all $j \in \mathcal{S}_{\bar{s}}$. If $\beta = \alpha^{-1}$, then*

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

b) $M^{\infty} = \imath(p\imath)^{-1}p.$

Proof.

Proof of part a): By Theorem 5.18, 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 $pM = p$. By

Figure 5.2: The Effect of Reduction on a Markov Graph

$$M$$

$$\begin{pmatrix} 0 & 1 & 0 & \frac{1}{6} \\ 1 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

$$\widehat{M}$$

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{1}{6} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

$$G\,(M)$$

$$G\left(\widehat{M}\right)$$
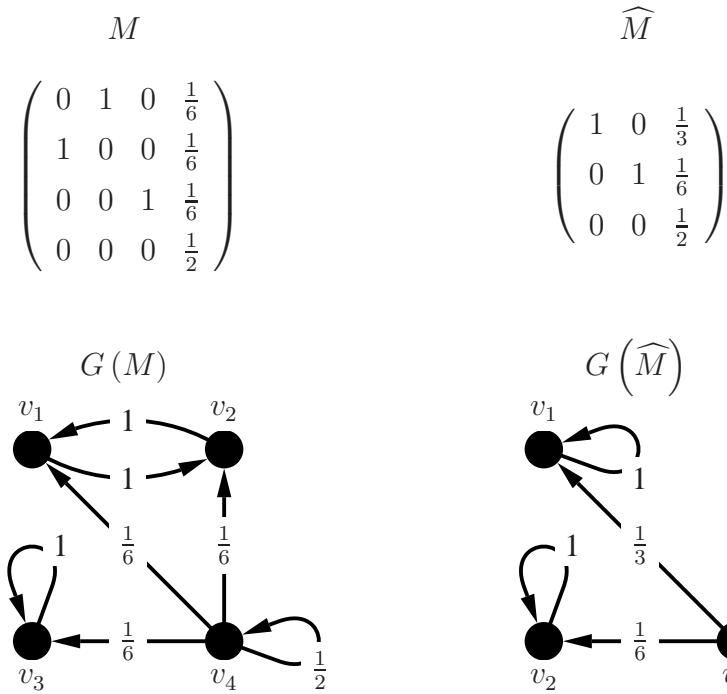
Theorem 5.6,

$$
\begin{aligned}
p\Lambda \;&=\; p(M - I) \\
&=\; \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t P_s \begin{pmatrix} M_{\overline{s},\overline{s}} & M_{\overline{s},s} \\ M_{s,\overline{s}} & M_{s,s} \end{pmatrix} P_s^t - p \\
&=\; \begin{pmatrix} M_{\overline{s},\overline{s}} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,\overline{s}} & M_{\overline{s},s} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} M_{s,s} \end{pmatrix} P_s^t - p \\
&=\; \begin{pmatrix} \widehat{M} & M_{\overline{s},s} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1}\left(\Lambda_{s,s} + I\right) \end{pmatrix} P_s^t - p \\
&=\; \begin{pmatrix} \widehat{M} & M_{\overline{s},s} - M_{\overline{s},s} - M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t - p \\
&=\; \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t - p \\
&=\; p - p \\
&=\; 0\,.
\end{aligned}
$$

In addition, we may show that the product, $p\iota$, is an invertible diagonal matrix. Again, using the notation of Theorem 5.6, $p\iota \;=\; \begin{pmatrix} I & -M_{\overline{s},s}\,\Lambda_{s,s}^{-1} \end{pmatrix} P_s^t P_s \begin{pmatrix} I \\ -\Lambda_{s,s}^{-1}\,M_{s,\overline{s}} \end{pmatrix} =$

$I + M_{\overline{s},s} \overline{\Lambda}^{-2} M_{s,\overline{s}}$. 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.

By Corollary 4.5, if $|s| = n - k$,

$$
\begin{aligned}
\left(M_{\overline{s},s} \Lambda_{s,s}^{-2} M_{s,\overline{s}}\right)_{i,j} &= \sum_{l=1}^{k} \left(-M_{\overline{s},s} \Lambda_{s,s}^{-1}\right)_{i,l} \left(-\Lambda_{s,s}^{-1} M_{s,\overline{s}}\right)_{l,j} \\
&= \sum_{l=1}^{k} \left[\sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_i,s_l)} W(M,\sigma)\right] \left[\sum_{\sigma' \in \mathcal{P}_M(s,s_l,\overline{s}_j)} W(M,\sigma')\right] \\
&= \sum_{l=1}^{k} \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_i,s_l)} \sum_{\sigma' \in \mathcal{P}_M(s,s_l,\overline{s}_j)} W(M,\sigma)W(M,\sigma') \\
&= \sum_{\sigma'' \in \mathcal{P}_M(s,s_i,\overline{s}_j)} W(M,\sigma'')
\end{aligned}
$$

where σ'' corresponds to the concatenation of paths, σ and σ'. Since $\overline{s}_j$ and $\overline{s}_i$ are in different closed classes, there are no no walks in $G(M)$ from $\overline{s}_j$ to $\overline{s}_i$, unless $i = j$. That is, $\mathcal{P}_M\left(s, s_i, \overline{s}_j\right) = \emptyset$ and the sum is 0, if $i \neq j$. Thus, $M_{\overline{s},s} \Lambda_{s,s}^{-2} M_{s,\overline{s}}$ is a non-negative, diagonal matrix. Moreover, 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, $(p\imath)_{j,j}^{-1} = \|\imath e_j\|_1^{-1}$.

Thus $\imath(p\imath)^{-1}$ is the result of dividing each column of $\imath$ by its corresponding column sum, so that the resulting columns are all distributions. By Theorem 5.18, v_j is a multiple of $\imath e_{\alpha(j)}$, so that $v_{\beta(j)}$ is a multiple of $\imath e_j$, and the jth column of $\imath(p\imath)^{-1}$ must be $v_{\beta(j)}$.

Proof of part b): Now we will now show that $\imath(p\imath)^{-1}p = M^{\infty}$ by appealing to Theorem 3.2. That is, we will show that $\imath(p\imath)^{-1}p = \pi_{\text{ker}}$. First, observe that, by Theorem 5.17 and Theorem 5.18, $\operatorname{im} \imath = \ker \Lambda$. By Lemma 3.1 c), 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, $\imath(p\imath)^{-1}pv = \pi_{\text{ker}}v$, so that $\imath(p\imath)^{-1}p = \pi_{\text{ker}} = M^{\infty}$.

$\square$

Example 5.22. Returning to Example 3.4 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 5.21. 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 $\square$

We can also justify our definition of equivalence.

Corollary 5.23. *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 $\ker(\Lambda_1) = \ker(\Lambda_2)$, 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. $\square$

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 natural 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, $\mu(\omega)$, 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 $\overline{\beta} \equiv \mathcal{S} - \beta$.

A chain, X_*, is *Markov* iff for all t, $s_0, \ldots, s_{t+1} \in \mathcal{S}$, 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 $\forall t$ s. t. $\Pr[X_t = s_t] > 0$, $\Pr[X_{t+1} = s_{t+1} \mid X_t = s_t]$ remains constant. Given an enumeration, $\iota : S_n \to \mathbb{R}$ of a superset of the state space, i.e., $\mathcal{S} \subset \mathrm{im}\,\iota$, we say that an $n \times n$ matrix, M, is a *transition matrix* of the chain *consistent with* ι iff $\Pr[X_{t+1} = \iota(i) \mid X_t = \iota(j)] = M_{i,j}$, whenever $\Pr[X_t = \iota(j)] > 0$. Notice that this implies that $n \geq |\mathcal{S}|$. If $n = |\mathcal{S}|$, so that $\mathcal{S} = \mathrm{im}\,\iota$, we say that M is a *minimal* transition matrix of the chain. The following Lemma shows that minimal transition matrices are essentially unique.

Lemma 5.24. *If M_1 and M_2 are two minimal transition matrices, consistent with ι_1 and ι_2, respectively, then $M_2 = P^t M_1 P$, where P is the permutation matrix such that $P_{i,j} =$*

---

[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.

$[\iota_1(i) = \iota_2(j)]$, *i.e.*, $P = P^\rho$ *for* $\rho = \iota_1^{-1}\iota_2$.

Proof. Since they are minimal, M_1 and M_2 are square matrices of the same size, namely, $n = |\mathcal{S}|$. For any $j \in S_n$, $\iota_2(j) \in \mathcal{S}$, so by definition, there is some $t \geq 0$ such that $\Pr[X_t = \iota_2(j)] > 0$. Therefore, for any $i \in S_n$,

$$
\begin{aligned}
\left(P^t M_1 P\right)_{i,j} &= \sum_{k,l=1}^{n} P_{i,k}^t (M_1)_{k,l} P_{l,j} = \sum_{k,l=1}^{n} P_{k,i} (M_1)_{k,l} P_{l,j} \\
&= \sum_{k,l=1}^{n} [\iota_1(k) = \iota_2(i)] \Pr[X_{t+1} = \iota_1(k) \mid X_t = \iota_1(l)] [\iota_1(l) = \iota_2(j)] \\
&= \Pr[X_{t+1} = \iota_2(i) \mid X_t = \iota_2(j)] = (M_2)_{i,j}
\end{aligned}
$$

Since the entries are equal, $M_2 = P^t M_1 P$. $\square$

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, ι, of a superset of its state space 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 5.25. *Given a Markov chain, X_*, and an enumeration, ι, with $\mathcal{S} \subset \text{im } \iota$, 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*
$$\left(\mathbb{E}\left[\vec{X}_t\right]\right)_j = \Pr[X_t = \iota(j)], \text{ for all } j.$$

Proof. By definition,

$$
\mathbb{E}\left[\vec{X}_t\right] = \sum_{k \in S_n} e_k \Pr\left[\vec{X}_t = e_k\right] = \sum_{k \in S_n} e_k \Pr[X_t = \iota(k)].
$$

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

Notice that the distribution of a stationary Markov chain consistent with a given Markov matrix, M, is determined by M and the initial distribution vector, $\mathbb{E}\left[\vec{X}_0\right]$, i.e., the distribution of X_0. This is a consequence of the following more general result.

Lemma 5.26. *For every sequence, $\sigma \in S_n(k)$,*

a) $\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_t = \iota\left(\sigma_k\right)\right] \neq 0$ *iff*

$$\Pr\left[X_{t-k} = \iota\left(\sigma_0\right)\right] \neq 0 \quad \text{and} \quad \sigma \in \mathcal{P}_M\left(k\right).$$

b) *More specifically,*

$$\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_t = \iota\left(\sigma_k\right)\right] = W\left(M, \sigma\right)\Pr\left[X_{t-k} = \iota\left(\sigma_0\right)\right] \qquad (5.10)$$

Proof. We prove both parts by induction on k. When $k = 0$, $W\left(M, \sigma\right) = 1$ and $\sigma \in \mathcal{P}_M\left(0\right)$, 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.

In the case when $\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] = 0$, we must have that

$$\Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_t = \iota\left(\sigma_k\right)\right] = 0.$$

Moreover, by induction, we know that either $\Pr\left[X_{t-k} = \iota\left(\sigma_0\right)\right] = 0$, in which case Equation 5.10 is trivially satisfied, or $\sigma' \notin \mathcal{P}_M\left(k - 1\right)$, in which case, $\sigma \notin \mathcal{P}_M\left(k\right)$, as well, so that $W\left(M, \sigma\right) = 0$, and Equation 5.10 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

$$\begin{aligned}
\Pr\big[X_{t-k} = \iota\left(\sigma_0\right), & \ldots, X_t = \iota\left(\sigma_k\right)\big] \\
&= \Pr\left[X_t = \iota\left(\sigma_k\right) \mid X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&\qquad \Pr\left[X_{t-k} = \iota\left(\sigma_0\right), \ldots, X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&= \Pr\left[X_t = \iota\left(\sigma_k\right) \mid X_{t-1} = \iota\left(\sigma_{k-1}\right)\right] \\
&\qquad W\left(M, \sigma'\right)\Pr\left[X_{t-k} = \iota\left(\sigma'_0\right)\right] \\
&= M_{\sigma_k, \sigma_{k-1}}W\left(M, \sigma'\right)\Pr\left[X_{t-k} = \iota\left(\sigma'_0\right)\right] \\
&= W\left(M, \sigma\right)\Pr\left[X_{t-k} = \iota\left(\sigma_0\right)\right]
\end{aligned}$$

where we also appeal to the definition of $W\left(M, \sigma\right)$ and the fact that $\sigma_0 = \sigma'_0$. $\square$

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 5.26 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.

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 5.26 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 5.26 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. To do this, however, we must first prove the following Lemma.

Lemma 5.27. *If M is a transition matrix of X_* with state space, $\mathcal{S}$, consistent with ι and $s = \iota^{-1}(\mathcal{S})$, then s is an invariant set of indices of M and $M_{s,s}$ is a minimal transition matrix of X_* consistent with $\iota' = \iota(s_*)$.*

Proof. The first part follows easily by contradiction. Assume that s is not invariant, so that there is some path, given by $\sigma \in \mathcal{P}_M(j, i, l)$ for $j \in s$ and $i \notin s$. Then, $W(M, \sigma) > 0$ and $\Pr[X_t = \iota(j)] > 0$ for some t. By Lemma 5.26 a), $\Pr[X_t = \iota(j), \ldots, X_{t+l} = \iota(i)] > 0$. In particular, we must have $\Pr[X_{t+l} = \iota(i)] > 0$, contradicting the assumption that $\iota(i) \notin \mathcal{S}$, i.e., $i \notin s$.

In particular, by Lemma 2.11 b), $M_{s,s}$ is Markov. If we let $\iota'(k) = \iota(s_k)$, $\iota' : S_{|s|} \to \mathbb{R}$ is a 1-1 mapping with $\text{im } \iota' = \iota(s) = \mathcal{S}$, i.e., an enumeration of $\mathcal{S}$. To verify that $M_{s,s}$ is a transition matrix of X_* consistent with ι', assume that $\Pr[X_t = \iota'(j)] > 0$. By definition, we then have $\Pr[X_t = \iota(s_j)] > 0$, so that $\Pr[X_{t+1} = \iota'(i) \mid X_t = \iota'(j)] = \Pr[X_{t+1} = \iota(s_i) \mid X_t = \iota(s_j)] = M_{s_i, s_j} = (M_{s,s})_{i,j}$, as needed. Since ι' enumerates $\mathcal{S}$, $M_{s,s}$ is minimal. $\square$

Therefore, we define the graph of X_*, $G(X_*) = G(M_{s,s})$.

We should point out that, although this definition appears to depend on the choice of transition matrix, it really only depends on the chain itself.

Theorem 5.28. *If M and M' are two transition matrices for X_*, consistent with ι and ι', respectively, with $s = \iota^{-1}(\mathcal{S})$ and $s' = \iota'^{-1}(\mathcal{S})$, then $G(M_{s,s}) = G\left(M'_{s',s'}\right)$.*

Proof. Since ι and ι' are both 1-1 and their images both contain $\mathcal{S}$, $|s| = |\mathcal{S}| = |s'| \equiv n$. Thus, there is a unique permutation of S_n, ρ, such that $\iota\left(s_{\rho(k)}\right) = \iota(s_k)$. By Lemma 5.27, $M_{s,s}$ and $M'_{s',s'}$ are both minimal transitions matrices for X_*, consistent with $\iota(s)$ and $\iota'(s')$, respectively. Thus, by Lemma 5.24, $M'_{s',s'} = (P^\rho)^t M_{s,s} P^\rho$, or $M'_{s'_i,s'_j} = M_{s_{\rho(i)},s_{\rho(i)}}$ for all $i,j \in S_n$.

We can now appeal to the definition of graph equality from section 1.1.1 with $G = G(M_{s,s}) = (V,E,s,t)$ and $G' = G\left(M'_{s',s'}\right) = (V',E',s',t')$. Explicitly, $V = V' = \{v_1,\ldots,v_n\}$, with $\delta(v_k) = v_{\rho(k)}$, which is clearly a 1-1 correspondence, since ρ is a permuation of S_n. Likewise, E and E' are the subsets of $V \times V$ corresponding to non-zero entries in $M_{s,s}$ and $M'_{s',s'}$, respectively, with $\gamma : E' \to E$ given by the restriction of $\delta \times \delta$ to E'. Since a Cartesian product of 1-1 correspondences is a 1-1 correspondence, we only need verify that $\delta \times \delta$ maps E' to E. By definition, $\alpha = (v_i, v_j)$ is in E' iff $0 \neq M'_{s'_i,s'_j} = M_{s_{\rho(i)},s_{\rho(i)}}$ iff $(v_{\rho}(i), v_{\rho}(i)) = \gamma(\alpha)$ is in E. Since s, s', t, t' are just the corresponding restrictions of the coordinate mappings, we clearly have $\delta s' = s\gamma$ and $\delta t' = t\gamma$, so that $G = G'$, as claimed. $\square$

Thus, we may define a set of states, β, of X_* to be strongly connected, closed, open, or transient iff the corresponding vertices in $G(X_*)$ are. These terms originally defined on graphs, now have compelling probabilistic interpretations. For example, a set of states is invariant iff the probability of ever transitioning away from the set is 0. To see this, observe that, by Lemma 5.26 a), $W(M, \sigma)$ represents the conditional probability of the random walk realizing the specific sequence of states corresponding to σ, given that $X_0 = \iota(\sigma_0)$. Using Lemma 1.1, we can also see that a state is transient iff there is a positive probability of transitioning away from it, since by Lemma 1.1 there exists a path to a state in a closed class, without ever returning (since there is no path out of a closed class).

Given a chain, X_*, and a Borel set, $\beta \in \mathcal{B}$, we want to define a new chain by reducing the time spent in a given set of states to 0, effectively eliminating them from consideration. We will see that this corresponds directly to the reduce construction of section 5.2. To make this precise, we first define the function $\gamma_{\beta,k} : \Omega \to \mathbb{N} \cup \{\infty\}$ so that $\gamma_{\beta,k}(\omega) = \min\{t > k \mid X_t(\omega) \notin \beta\}$, with the conventions that $\min \emptyset = \infty$ and $\gamma_{\beta,\infty} = \infty$. Notice that, by definition, $\gamma_{\beta,k} > k$, unless $k = \infty$. We can easily show that these are measureable

functions, i.e., random variables. More specifically, the events $\{\gamma_{\beta,k} = t\}$ can be expressed solely in terms of events of the form $\{X_i \in \beta_i\}$ for $i \leq t$ and $\beta_i \in \mathcal{B}$. Thus, these functions are *Markov times* (Iosifescu, 1980).

Lemma 5.29. *For any set of states, β, of a finite-state, stationary Markov process,*

$$\{\gamma_{\beta,k} = t\} = \begin{cases} \{X_t \in \overline{\beta}\} \cap \bigcap_{k < t' < t} \{X_{t'} \in \beta\} & \text{if } k < t < \infty \\ \bigcap_{k < t'} \{X_{t'} \in \beta\} & \text{if } t = \infty \\ \emptyset & \text{otherwise} \end{cases} \quad (5.11)$$

so that $\gamma_{\beta,k}$ is a Markov time.

Proof. Since $\mathcal{S}$ is a finite set, and $\beta, \overline{\beta} \subset \mathcal{S}$, $\beta, \overline{\beta} \in \mathcal{B}$. Now observe that $\omega \in \{\gamma_{\beta,k} = t\}$ iff $\gamma_{\beta,k}(\omega) = t$ iff $t = \min \{t' > k \mid X_{t'}(\omega) \notin \beta\}$. By definition, for $\omega \in \{\gamma_{\beta,k} = t\}$,

$$t = \infty \iff \{t' > k \mid X_{t'}(\omega) \notin \beta\} = \emptyset \iff X_{t'}(\omega) \in \beta, \forall\, t' > k$$
$$\iff \omega \in \bigcap_{k < t'} \{X_{t'} \in \beta\}.$$

Otherwise, $t < \infty$ so that $t \in \{t'' > k \mid X_{t''}(\omega) \notin \beta\}$ with $k < t < \infty$, $X_t(\omega) \in \overline{\beta}$, and $t' \notin \{t'' > k \mid X_{t''}(\omega) \notin \beta\}$, i.e., $X_{t'}(\omega) \in \beta$, for every $k < t' < t$. In other words, $\omega \in \{X_t \in \overline{\beta}\} \cap \bigcap_{k < t' < t} \{X_{t'} \in \beta\}$. Thus, Equation 5.11 holds. $\square$

We now prove that $\gamma_{\beta,k}$ is almost always finite, when β is open. Such a random variable is known as a *stopping time*. Specifically, $\gamma_{\beta,k}$ is the first *hitting time* for $\overline{\beta}$ greater than k. In fact, this effectively characterizes open sets of states.

Lemma 5.30. *A set of states, β, of a finite-state, stationary Markov process, X_*, is open iff* $\Pr\left[\gamma_{\beta,k} = \infty\right] = 0, \forall\, k.$

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 4.4, the uth column sum of $M_{s,s}^q$ is

$$\begin{aligned} JM_{s,s}^q e_u &= \sum_{v \in S_{|s|}} \left(M_{s,s}^q\right)_{v,u} = \sum_{v \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_{M_{s,s}}(v,u,q)} W\left(M_{s,s}, \sigma\right) \\ &= \sum_{v \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_M(s, s_v, s_u, q)} W\left(M, \sigma\right) \end{aligned} \quad (5.12)$$

Moreover,

$$\Pr\left[\bigcap_{r=j}^{j+q}\{X_r \in \beta\}\right] = \Pr\left[X_j \in \beta, \ldots, X_{j+q} \in \beta\right]$$

$$= \sum_{i_0,\ldots,i_q \in s} \Pr\left[X_j = \iota\left(i_0\right), \ldots, X_{j+q} = \iota\left(i_q\right)\right]$$

$$= \sum_{u,v \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_M(s,s_v,s_u,q)} \Pr\left[X_j = \iota\left(\sigma_0\right), \ldots, X_{j+q} = \iota\left(\sigma_q\right)\right]$$

Therefore,

$$\Pr\left[\bigcap_{r=j}^{j+q}\{X_r \in \beta\}\right]$$

$$= \sum_{u,v \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_M(s,s_v,s_u,q)} W\left(M,\sigma\right)\Pr\left[X_j = \iota\left(\sigma_0\right)\right] \quad \text{by Equation 5.10}$$

$$= \sum_{u \in S_{|s|}} JM_{s,s}^q e_u \Pr\left[X_j = \iota\left(s_u\right)\right] \quad \text{by Equation 5.12}$$

$$= JM_{s,s}^q \sum_{u \in S_{|s|}} e_u \Pr\left[X_j = \iota\left(s_u\right)\right]$$

In particular, by Lemma 5.29,

$$\Pr\left[\gamma_{\beta,k} = \infty\right] = \lim_{q\to\infty} \Pr\left[\bigcap_{r=k}^{k+q}\{X_r \in \beta\}\right] = \lim_{q\to\infty} JM_{s,s}^q \sum_{u \in S_{|s|}} e_u \Pr\left[X_j = \iota\left(s_u\right)\right]$$

If β is open, by Theorem 4.4 a), $\lim_{q\to\infty} JM_{s,s}^q = 0$ so that $\Pr\left[\gamma_{\beta,k} = \infty\right] = 0$.

Conversely, assume that $\Pr\left[\gamma_{\beta,k} = \infty\right] = 0$, $\forall k$. 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,

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

Since $0 = \lim_{q\to\infty} \Pr\left[\bigcap_{r=k}^{k+q}\{X_r \in \beta\}\right]$, $0 = \Pr\left[X_j = \iota\left(s_u\right)\right] \lim_{q\to\infty} JM_{s,s}^q e_u$. Because $\Pr\left[X_j = \iota(s_u)\right] \neq 0$, we must have $\lim_{q\to\infty} JM_{s,s}^q e_u = 0$. In particular,

$$0 = \sum_j \lim_{q\to\infty} JM_{s,s}^q e_j = \sum_{i,j} \lim_{q\to\infty} e_i^t M_{s,s}^q e_j \geq \lim_{q\to\infty} \left(M_{s,s}^q\right)_{i,j} \geq 0$$

so that $\lim_{q\to\infty} \left(M_{s,s}^q\right)_{i,j} = 0$ and $\lim_{q\to\infty} M_{s,s}^q = 0$. Appealing to the contrapositive of Theorem 4.4 d), we may conclude that β is open. $\square$

Using $\gamma_{\beta,k}$, we may define $\tau_{\beta,k}$ inductively so that $\tau_{\beta,0} = \gamma_{\beta,-1}$ and $\tau_{\beta,k} = \gamma_{\beta,\tau_{\beta,k-1}}$ for $k > 0$. This reduces the time take to pass through any state in β to 0, effectively "eliminating" β from consideration.

Lemma 5.31. *For any set of states, β, of a finite-state, stationary Markov process, if $k > 0$,*

$$\left\{\tau_{\beta,k} = t\right\} = \begin{cases} \bigcup_{t' < t} \left\{\gamma_{\beta,t'} = t\right\} \cap \left\{\tau_{\beta,k-1} = t'\right\} & \text{if } t < \infty \\ \bigcup_{t'} \left\{\gamma_{\beta,t'} = \infty\right\} \cap \left\{\tau_{\beta,k-1} = t'\right\} & \text{if } t = \infty \end{cases} \tag{5.13}$$

In particular, $\tau_{\beta,k}$ is a Markov time. If $\tau_{\beta,k}(\omega) < \infty$, $\forall k$, $\left\{\tau_{\beta,k}(\omega)\right\}_{k=0}^{\infty}$ is a strictly increasing sequence. If β is open, $\tau_{\beta,k} < \infty$ with probability 1, i.e., it is a stopping time.

Proof. By definition, $\tau_{\beta,k+1}(\omega) = \gamma_{\beta,\tau_{\beta,k}}$, i.e., for $t = \tau_{\beta,k+1}(\omega)$ and $t' = \tau_{\beta,k}(\omega)$, $t = \gamma_{\beta,t'}$. By definition of γ, we must have $t' < t$, unless $t = t' = \infty$. In particular, $\tau_{\beta,k}(\omega) \leq \tau_{\beta,k+1}(\omega)$, with inequality unless $\tau_{\beta,k}(\omega) = \infty$.

Therefore, for $k > 0$, $\omega \in \left\{\tau_{\beta,k} = t\right\}$ iff $t = \gamma_{\beta,t'}(\omega)$ and $t' = \tau_{\beta,k-1}(\omega)$, where either $t' < t < \infty$, or $t' \leq t = \infty$. Using set notation, this becomes $\omega \in \bigcup_{t' < t} \left\{\gamma_{\beta,t'} = t\right\} \cap \left\{\tau_{\beta,k-1} = t'\right\}$, when $t < \infty$ and $\omega \in \bigcup_{t'} \left\{\gamma_{\beta,t'} = \infty\right\} \cap \left\{\tau_{\beta,k-1} = t'\right\}$, when $t = \infty$. Thus, Equation 5.13 holds.

By Lemma 5.29, $\left\{\gamma_{\beta,t'} = t\right\}$ is measureable for $t, t' \leq \infty$, therefore, by induction, so is $\left\{\tau_{\beta,k} = t\right\}$. In other words, $\tau_{\beta,k}$ is a Markov time. From our ealier observations, $\tau_{\beta,k} \leq \tau_{\beta,k+1}$ for all k, except when $\tau_{\beta,k} = \infty$. That is, $\tau_{\beta,k}$ is strictly increasing, when $\tau_{\beta,k}(\omega) < \infty$, $\forall k$.

If β is open, by Lemma 5.30, $\Pr\left[\gamma_{\beta,t'} = \infty\right] = 0$. Therefore,

$$\Pr\left[\tau_{\beta,k} = \infty\right] = \Pr\left[\bigcup_{t'} \left\{\gamma_{\beta,t'} = \infty\right\} \cap \left\{\tau_{\beta,k-1} = t'\right\}\right]$$

$$\leq \sum_{t'} \Pr\left[\gamma_{\beta,t'} = \infty \text{ and } \tau_{\beta,k-1} = t'\right] \leq \sum_{t'} \Pr\left[\gamma_{\beta,t'} = \infty\right] = 0$$

and $\tau_{\beta,k}$ is a stopping time. $\square$

We may now give the following intuitive description of $\tau_{\beta,k}$.

Lemma 5.32. *If $\tau_{\beta,k}(\omega) < \infty$, $\forall k$,*

a) $\tau_{\beta,k}(\omega) = \min\left\{t \geq 0 \mid k+1 = \left|\{0 \leq t' \leq t \mid X_{t'}(\omega) \in \overline{\beta}\}\right|\right\}.$

b) $\left\{\tau_{\beta,k}(\omega)\right\}_{k=0}^{\infty}$ *is an increasing enumeration of*

$$T(\omega) = \left\{t \geq 0 \mid X_t(\omega) \in \overline{\beta}\right\}.$$

In particular, if β is open, a) and b) hold with probability 1.

Proof. For convenience, we will drop the notation for evaluation, since all random variables (i.e., X_t, $\gamma_{\beta,k}$, and $\tau_{\beta,k}$, etc.) will always be evaluated at a fixed value, ω, for which $\tau_{\beta,k}(\omega) < \infty$. We begin by proving part b). By definition, $\tau_{\beta,k} = \gamma_{\beta,t'}$ for some t'. Moreover, if $\tau_{\beta,k} < \infty$, then $\gamma_{\beta,t'} < \infty$, and by definition, we must have $\gamma_{\beta,t'} \in T$, so that $\tau_{\beta,k} \in T$, as well. By Lemma 5.31, $\tau_{\beta,k}$ is a strictly increasing increasing sequence. In particular, it is 1-1 as a function of k.

It remains to observe that this function, which we will denote as $\tau_{\beta,*}$, maps onto T. We prove this by contradiction, so assume $T \setminus \operatorname{im}\tau_{\beta,*} \neq \emptyset$ and take $t = \min\left(T \setminus \operatorname{im}\tau_{\beta,*}\right)$. Thus, either $t' \notin T$ for all $t' < t$, or for some k and $t' < t$, $\tau_{\beta,k} = t' \in T$. In the former case, $X_t \in \overline{\beta}$ and $X_{t'} \in \beta$ for all $t' < t$. Thus, $t = \min\left\{t' > -1 \mid X_{t'} \notin \beta\right\}$, so that $t = \gamma_{\beta,-1} = \tau_{\beta,0}$ contradicting the assumption that $t \notin \operatorname{im}\tau_{\beta,*}$.

Now assume that $\tau_{\beta,k} = t' \in T$ for some k and $t' < t$, and take the largest such t'. This means that for $t' < t'' < t$, $t'' \notin \operatorname{im}\tau_{\beta,*}$, so that we must have $t'' \notin T$ or $X_{t''} \in \beta$. In other words, $t = \min\left\{t'' > t' \mid X_{t''} \notin \beta\right\}$, so that $t = \gamma_{\beta,t'} = \gamma_{\beta,\tau_{\beta,k}} = \tau_{\beta,k+1}$, again contradicting the assumption that $t \notin \operatorname{im}\tau_{\beta,*}$. Thus, $T \setminus \operatorname{im}\tau_{\beta,*} = \emptyset$ and $T = \operatorname{im}\tau_{\beta,*}$. In other words, $\tau_{\beta,*}$ is an increasing enumeration of $T = \left\{t \geq 0 \mid X_t \in \overline{\beta}\right\}$, and we have proven part b).

Part a) now follows easily. Since $\tau_{\beta,*}$ is an enumeration of $T = \left\{t \geq 0 \mid X_t \in \overline{\beta}\right\}$, $\left\{0 \leq t' \leq t \mid X_{t'} \in \overline{\beta}\right\} = \left\{0 \leq t' \leq t \mid t' = \tau_{\beta,k'} \text{ for some } k'\right\}$ and

$$\left|\{0 \leq t' \leq t \mid X_{t'} \in \overline{\beta}\}\right| = \left|\left\{k' \mid \tau_{\beta,k'} \leq t\right\}\right|.$$

Since $\tau_{\beta,k}$ is an increasing sequence starting from $k = 0$,

$$\left|\left\{k' \mid \tau_{\beta,k'} \leq t\right\}\right| = 1 + \max_{k'}\left\{k' \mid \tau_{\beta,k'} \leq t\right\}.$$

In particular,

$$k+1 = \left|\{0 \leq t' \leq t \mid X_{t'} \in \overline{\beta}\}\right| \iff k = \max_{k'}\left\{k' \mid \tau_{\beta,k'} \leq t\right\}.$$

Therefore,

$$\min \left\{ t \geq 0 \mid k+1 = \left| \{0 \leq t' \leq t \mid X_{t'} \in \overline{\beta} \} \right| \right\}$$
$$= \min \left\{ t \geq 0 \mid k = \max_{k'} \left\{ k' \mid \tau_{\beta,k'} \leq t \right\} \right\}$$
$$= \tau_{\beta,k}.$$

Finally, if β is open, by Lemma 5.31,

$$\Pr \left[\bigcup_k \{ \tau_{\beta,k} = \infty \} \right] = \sum_k \Pr \left[\tau_{\beta,k} = \infty \right] = \sum_k 0 = 0$$

Therefore, $\tau_{\beta,k} < \infty$, $\forall \, k$, so that a) and b) hold, with probability 1. $\square$

Using the language of Markov chain theory, Lemma 5.32 a) says that $\tau_{\beta,k}$ is the $k+1^{\text{st}}$ "hitting time" for $\overline{\beta}$.

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 reduce 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.

Using Lemma 5.32, we may show that the reduction operator is "natural" in the sense that it behaves as expected under iteration. Lemma 5.32 b) implies that, with probability 1, $\pi_{\beta,*}(X_*)$ is simply the result of deleting those entries of X_* with values in β. 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 5.33. *Given a Markov chain, X_*, and open Borel sets, β_1, β_2, and $\beta = \beta_1 \cup \beta_2$, $\pi_{\beta,*}(X_*) = \pi_{\beta_1,*}\left(\pi_{\beta_2,*}(X_*)\right)$ with probability 1.*

Proof. Let $X_*^2 = \pi_{\beta_2,*}(X_*)$, $T_2 = \{t \geq 0 \mid X_t \in \overline{\beta_2}\}$, $T_1 = \{t \geq 0 \mid X_t^2 \in \overline{\beta_1}\}$, and $T = \{t \geq 0 \mid X_t \in \overline{\beta}\}$. Let $\Omega' = \{\tau_{\beta_2,k}, \tau_{\beta_1,k}, \tau_{\beta,k} < \infty, \forall \, k \geq 0\}$. By Lemma 5.31, $\Pr \Omega' = 1$, and, by Lemma 5.32 b), $\tau_{\beta_2,*}$, $\tau_{\beta_1,*}$, and $\tau_{\beta,*}$ are increasing enumerations of T_2, T_1, and T, respectively, on Ω'.

As before, for the remainder of the proof, we will restrict attention to an fixed, arbitrary $\omega \in \Omega'$ and omit the notation for evaluation at ω from all random variables. In this case, we have the following chain of equivalences:

$\exists\, k \geq 0,\ t = \tau_{\beta_2, \tau_{\beta_1, k}}$

$$\iff \exists\, k, t' \geq 0,\ t' = \tau_{\beta_1, k} \text{ and } t = \tau_{\beta_2, t'}$$

$$\iff \exists\, t' \in T_1,\ t = \tau_{\beta_2, t'} \qquad\qquad\qquad \text{since } \tau_{\beta_1, *} \text{ enumerates } T_1$$

$$\iff \exists\, t' \geq 0,\ X_{t'}^2 \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2, t'} \qquad \text{by definition of } T_1$$

$$\iff \exists\, t' \geq 0,\ X_{\tau_{\beta_2, t'}} \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2, t'} \quad \text{by definition of } X_*^2$$

$$\iff \exists\, t' \geq 0,\ X_t \in \overline{\beta_1} \text{ and } t = \tau_{\beta_2, t'}$$

$$\iff X_t \in \overline{\beta_1} \text{ and } t \in T_2 \qquad\qquad\qquad \text{since } \tau_{\beta_2, *} \text{ enumerates } T_2$$

$$\iff X_t \in \overline{\beta_1} \text{ and } X_t \in \overline{\beta_2} \qquad\qquad\qquad \text{by definition of } T_2$$

$$\iff X_t \in \overline{\beta_1} \cap \overline{\beta_2} = \overline{\beta_1 \cup \beta_2} = \overline{\beta}$$

$$\iff t \in T \qquad\qquad\qquad\qquad\qquad\qquad \text{by definition of } T$$

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 strictly increasing mappings, so is their composite. In particular, $\tau_{\beta_2, \tau_{\beta_1, t}}$ is an increasing enumeration of T. Since increasing enumerations are unique, $\tau_{\beta_2, \tau_{\beta_1, t}} = \tau_{\beta, t}$, and $\pi_{\beta_1, *}\left(\pi_{\beta_2, *}\left(X_*\right)\right) = \pi_{\beta_1, *}\left(X_*^2\right) = X_{\tau_{\beta_1, *}}^2 = X_{\tau_{\beta_2, \tau_{\beta_1, t}}} = X_{\tau_{\beta, t}} = \pi_{\beta, *}\left(X_*\right)$ on Ω', i.e., with probability 1. $\square$

We now wish to identify the transition matrix for $\pi_{\beta, *}\left(X_*\right)$. 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

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

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, $\mathcal{S}_n\left(s, i, j, l, 0\right) = \mathcal{S}_n\left(s, i, j, l\right)$.

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

Theorem 5.34. *If X_* is a finite state, stationary Markov chain with state space, $\mathcal{S}$, which is ι-consistent with a $n \times n$ transition matrix, M, β is open, and $\widetilde{X}_* = \pi_{\beta, *}\left(X_*\right)$, then $\overline{\beta}$ is the*

state space for $\widetilde{X}_*$ and $\iota'(k) = \iota\left(\overline{s}_k\right)$ enumerates a superset of $\overline{\beta}$, where $s = \iota^{-1}(\beta) \subset S_n$. For $t > 0$,

$$\Pr\left[\widetilde{X}_t = \iota'(k)\right] = \sum_{l=0}^{\infty}\sum_{j\in s}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t)} W(M,\sigma)\Pr\left[X_0 = \iota(j)\right]$$

$$+ \sum_{l=0}^{\infty}\sum_{j\in\overline{s}}\sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t-1)} W(M,\sigma)\Pr\left[X_0 = \iota(j)\right] \quad (5.14)$$

and

$$\Pr\left[\widetilde{X}_0 = \iota'(k)\right] = \sum_j p_{k,j}\Pr\left[X_0 = \iota(j)\right] \quad (5.15)$$

where $\left(\widehat{M}, p, \iota\right)$ 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 this distribution as an column vector, as in Lemma 5.25.

Proof. Notice that since ι and s_* are 1-1 functions, so is $\iota(s_*)$. By definition, $\operatorname{im}\iota(s_*) = \iota(s) = \beta\cap\mathcal{S}$, so that $\iota(s_*)$ enumerates $\beta\cap\mathcal{S}$. Similarly, $\iota' = \iota(\overline{s}_*)$ enumerates $\iota(\overline{s}) = \iota(S_n - s) = \operatorname{im}\iota - \beta$, which is a superset of $\mathcal{S} - \beta = \overline{\beta}$.

By definition, $\pi_{\beta,t}(X_*)(\omega) \in \overline{\beta}$ for all t and ω. In particular, $\widetilde{X}_t = x$ only if $x \in \overline{\beta}$ iff $x = \iota(\overline{s}_k) = \iota'(k)$ for some $k \in S_{|\overline{s}|}$. Now remember that, when $\tau_{\beta,t} < \infty$, $\{\tau_{\beta,j}\}_{j=0}^t$ is an increasing sequence of integers. In particular, $\tau_{\beta,t} \geq t$, $l \equiv \tau_{\beta,t} - t \geq 0$, and dividing into

cases gives

$$\Pr\left[\widetilde{X}_t = \iota'(k)\right] = \Pr\left[\widetilde{X}_t = \iota'(k), \tau_{\beta,t} < \infty\right] \qquad \text{by Lemma 5.31}$$

$$= \Pr\left[X_{\tau_{\beta,t}} = \iota(\overline{s}_k)\right] \qquad \text{by defn of } \widetilde{X}_t$$

$$= \sum_{l=0}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \qquad \text{by cases}$$

$$= [t = 0]\Pr\left[X_t = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right]$$

$$+ \sum_{l=[t=0]}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \qquad \text{by notational trick}$$

$$= [t = 0]\Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right]$$

$$+ \sum_{l=[t=0]}^{\infty} \Pr\left[X_{l+t} = \iota(\overline{s}_k), l = \tau_{\beta,t} - t\right] \qquad \text{by defn of } \tau_{\beta,t}$$

$$= [t = 0]\Pr\left[X_0 = \iota(\overline{s}_k), \tau_{\beta,0} = 0\right]$$

$$+ \sum_{l=[t=0]}^{\infty} \sum_{j=1}^{n} \Pr\left[X_{l+t} = \iota(\overline{s}_k), X_0 = \iota(j), l = \tau_{\beta,t} - t\right] \quad \text{by cases}$$

The first time, cases were based on $l = \tau_{\beta,t} - t$, and the second time according to the initial state of the process, using the fact that ι enumerates a superset of the state space of X_*. In between, 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 $l + t > 0$. Therefore, after splitting the summation according to the initial state, we can continue to

divide into cases based on the intermediate states of X_*.

$$
\begin{aligned}
\Pr\left[\widetilde{X}_t = \iota'(k)\right] \;=\; & [t=0]\Pr\left[X_0 = \iota\left(\overline{s}_k\right), \tau_{\beta,0} = 0\right] \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in\overline{s}} \Pr\left[X_{l+t} = \iota\left(\overline{s}_k\right), X_0 = \iota(j), \text{ and } l = \tau_{\beta,t} - t\right] \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in s} \Pr\left[X_{l+t} = \iota\left(\overline{s}_k\right), X_0 = \iota(j), \text{ and } l = \tau_{\beta,t} - t\right] \\
\;=\; & [t=0]\Pr\left[X_0 = \iota\left(\overline{s}_k\right), \tau_{\beta,0} = 0\right] \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in s} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l+t)} \Pr\left[X_i = \iota\left(\sigma_i\right), 0\leq i\leq l+t\right] \quad (5.16) \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in\overline{s}} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l+t)} \Pr\left[X_i = \iota\left(\sigma_i\right), 0\leq i\leq l+t\right] \\
\;=\; & [t=0]\Pr\left[X_0 = \iota\left(\overline{s}_k\right), \tau_{\beta,0} = 0\right] \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in s} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t)} \Pr\left[X_i = \iota\left(\sigma_i\right), 0\leq i\leq l+t\right] \quad (5.17) \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in\overline{s}} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t-1)} \Pr\left[X_i = \iota\left(\sigma_i\right), 0\leq i\leq l+t\right] \\
\;=\; & [t=0]\Pr\left[X_0 = \iota\left(\overline{s}_k\right), \tau_{\beta,0} = 0\right] \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in s} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t)} W(M,\sigma)\Pr\left[X_0 = \iota(j)\right] \quad (5.18) \\
& + \sum_{l=[t=0]}^{\infty} \sum_{j\in\overline{s}} \sum_{\sigma\in S_n(s,\overline{s}_k,j,l,t-1)} W(M,\sigma)\Pr\left[X_0 = \iota(j)\right]
\end{aligned}
$$

In Equation 5.16, we used the fact that $\mathrm{im}\,\iota$ is a superset of $\mathcal{S}$. Equation 5.17 follows from the fact that, in the context of any of the summation terms, $X_0 = \iota(j)$, $X_{l+t} = \iota\left(\overline{s}_k\right)$, $l = \tau_{\beta,t} - t$, and $l + t > 0$. Therefore, by Lemma 5.32 a), $\{X_i\}_{i=0}^{l+t}$ takes exactly $t+1$ values in $\overline{\beta}$ and l values in β, with probability 1, so that $\sigma \in S_n(s, \overline{s}_k, j, l, t)$, if $j \in s$, and $\sigma \in S_n(s, \overline{s}_k, j, l, t-1)$, if $j \notin s$. Equation 5.18 follows by Lemma 5.26 b).

Equation 5.18 simplifies to Equation 5.14, 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 5.18 is non-zero, so that $\Pr\left[\widetilde{X}_t = \iota'(k)\right] > 0$.

Otherwise, by Lemma 5.26, and the fact that ι enumerates a superset of the state space of X_*,

$$\Pr\left[X_t = \iota'(k)\right] = \sum_{j \in S_n} \Pr\left[X_t = \iota\left(\overline{s}_k\right), X_0 = \iota\left(j\right)\right]$$

$$= \sum_{j \in S_n} \sum_{\sigma \in S_n(\overline{s}_k, j, t)} \Pr\left[X_i = \iota\left(\sigma_i\right), i = 0, \ldots, t\right]$$

$$= \sum_{j \in S_n} \sum_{\sigma \in S_n(\overline{s}_k, j, t)} W\left(M, \sigma\right) \Pr\left[X_0 = \iota\left(j\right)\right]$$

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 $\Pr\left[X_0 = \iota\left(j\right)\right]$ and $W\left(M, \sigma\right)$ are non-zero. 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 5.14 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 5.15, take $t = 0$. Since $S_n(s, \overline{s}_k, j, l, t-1) = \emptyset$, Equation 5.18 simplifies to

$$\Pr\left[\widetilde{X}_0 = \iota'(k)\right] = \Pr\left[X_0 = \iota\left(\overline{s}_k\right), \tau_{\beta,0} = 0\right]$$

$$+ \sum_{l=1}^{\infty} \sum_{j \in s} \sum_{\sigma \in S_n(s, \overline{s}_k, j, l, 0)} W(M, \sigma) \Pr\left[X_0 = \iota\left(j\right)\right]$$

$$= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{l=1}^{\infty} \sum_{j \in s} \sum_{\sigma \in S_n(s, \overline{s}_k, j, l)} W(M, \sigma) \Pr\left[X_0 = \iota\left(j\right)\right]$$

$$= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{j \in s} \sum_{l=1}^{\infty} \sum_{\sigma \in S_n(s, \overline{s}_k, j, l)} W(M, \sigma) \Pr\left[X_0 = \iota\left(j\right)\right]$$

$$= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{j \in s} \sum_{\sigma \in S_n(s, \overline{s}_k, j)} W(M, \sigma) \Pr\left[X_0 = \iota\left(j\right)\right]$$

$$= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{j \in s} \sum_{\sigma \in \mathcal{P}_M(s, \overline{s}_k, j)} W(M, \sigma) \Pr\left[X_0 = \iota\left(j\right)\right]$$

$$= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{j \in S_{|s|}} \sum_{\sigma \in \mathcal{P}_M(s, \overline{s}_k, s_j)} W(M, \sigma) \Pr\left[X_0 = \iota\left(s_j\right)\right]$$

Now by Theorem 5.8 a), right-hand side of Equation 5.15 as follows,

$$
\begin{aligned}
\sum_j p_{k,j} \Pr\left[X_0 = \iota(j)\right] &= \sum_j e_k^t p e_j \Pr\left[X_0 = \iota(j)\right] \\
&= \sum_j \left(e_{\overline{s}_k}^t + \sum_{l=1}^{|s|} \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_k,s_l)} W(M,\sigma) e_{s_l}^t \right) e_j \Pr\left[X_0 = \iota(j)\right] \\
&= \sum_j e_{\overline{s}_k}^t e_j \Pr\left[X_0 = \iota(j)\right] \\
&\quad + \sum_j \sum_{l=1}^{|s|} \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_k,s_l)} W(M,\sigma) e_{s_l}^t e_j \Pr\left[X_0 = \iota(j)\right] \\
&= \sum_j [\overline{s}_k = j] \Pr\left[X_0 = \iota(j)\right] \\
&\quad + \sum_j \sum_{l=1}^{|s|} \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_k,s_l)} W(M,\sigma)\,[s_l = j] \Pr\left[X_0 = \iota(j)\right] \\
&= \Pr\left[X_0 = \iota\left(\overline{s}_k\right)\right] + \sum_{l=1}^{|s|} \sum_{\sigma \in \mathcal{P}_M(s,\overline{s}_k,s_l)} W(M,\sigma) \Pr\left[X_0 = \iota\left(s_l\right)\right]
\end{aligned}
$$

which, up to a change of summation index, is the same result as above, so that

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

□

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 reduce construction of section 5.2 to the transition matrix for X_*. The proof is similar to that of Theorem 5.34, 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 5.35. *If X_* is a finite state, stationary Markov chain which is ι-consistent with transition matrix, M, and β is an open set of states, 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 $\iota'(j) = \iota(\overline{s}_j)$. If M is minimal, then so is $\widehat{M}$.*

Proof. The proof is similar to that of Theorem 5.34. As before, define $\widetilde{X}_t = \pi_{\beta,t}(X_*)$, and notice that $\iota(s_*)$ enumerates β, while $\iota(\overline{s}_*)$ enumerates a superset of $\overline{\beta}$. By Lemma 5.31, we may restrict attention to the case, $\tau_{\beta,t} < \tau_{\beta,t+1} < \infty$ and again exploit the fact that $t \leq \tau_{\beta,t}$.

$$\Pr\left[\widetilde{X}_{t+1} = \iota'(k), \widetilde{X}_t = \iota'(k')\right] = \Pr\left[\widetilde{X}_{t+1} = \iota'(k), \widetilde{X}_t = \iota'(k'), \tau_{\beta,t+1} < \infty\right]$$

$$= \sum_{m=0}^{\infty}\sum_{l=0}^{\infty} \Pr\left[X_{l+m+t+1} = \iota(\overline{s}_k), X_{m+t} = \iota(\overline{s}_{k'}),\right.$$

$$\left. l = \tau_{\beta,t+1} - m - t - 1, m = \tau_{\beta,t} - t\right]$$

Assuming that $X_{m+t} = \iota(\overline{s}_{k'})$ and $m = \tau_{\beta,t} - t$, Lemma 5.32 a) implies $X_{l+m+t+1} = \iota(\overline{s}_k)$ with $l = \tau_{\beta,t+1} - m - t - 1$ iff $X_i \in \beta$ for $m+t < i \leq l+m+t$ with $X_{l+m+t+1} = \iota(\overline{s}_k)$. Thus,

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

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

Appealing again to Lemma 5.32 a), we see that $X_{m+t} = \iota(\overline{s}_{k'})$ 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 $X_i = \iota(\sigma''_i)$ for $0 \leq i \leq m+t$, for some $\sigma'' \in S'_n(s, \overline{s}_k, m, t)$. Thus, for $\sigma' \in \mathcal{S}_n(s, \overline{s}_k, \overline{s}_{k'}, l+1)$,

$$\Pr\left[X_{m+t+i} = \iota(\sigma'_i), i = 0,\ldots,l+1, m = \tau_{\beta,t} - t\right]$$

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

Therefore,

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

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

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

Now appealing to Lemma 5.26 b), we have

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

$$= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \sum_{\sigma' \in \mathcal{S}_n\left(s, \overline{s}_k, \overline{s}_{k'}, l+1\right)} \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W(M, \sigma' * \sigma'') \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \sum_{\sigma' \in \mathcal{S}_n\left(s, \overline{s}_k, \overline{s}_{k'}, l+1\right)} \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma'\right) W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \sum_{\sigma' \in \mathcal{S}_n\left(s, \overline{s}_k, \overline{s}_{k'}, l+1\right)} W\left(M, \sigma'\right) \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

We may then exploit Theorem 5.8 as follows,

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

$$= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \sum_{\sigma' \in \mathcal{S}_n\left(s, \overline{s}_k, \overline{s}_{k'}, l+1\right)} W\left(M, \sigma'\right) \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \sum_{m=0}^{\infty} \sum_{\sigma' \in \mathcal{S}_n\left(s, \overline{s}_k, \overline{s}_{k'}\right)} W\left(M, \sigma'\right) \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \sum_{m=0}^{\infty} \sum_{\sigma' \in \mathcal{P}_M\left(s, \overline{s}_k, \overline{s}_{k'}\right)} W\left(M, \sigma'\right) \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \sum_{m=0}^{\infty} \widehat{M}_{k,k'} \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

$$= \widehat{M}_{k,k'} \sum_{m=0}^{\infty} \sum_{\sigma'' \in S_n'\left(s, \overline{s}_{k'}, m, t\right)} W\left(M, \sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right]$$

---

[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 $(3, 1, 4, 4, 7, 2)$.

Reversing our previous calculations gives

$$\sum_{m=0}^{\infty} \sum_{\sigma'' \in S_n'(s,\bar{s}_{k'},m,t)} W\left(M,\sigma''\right) \Pr\left[X_0 = \iota\left(\sigma_0''\right)\right] =$$

$$= \sum_{m=0}^{\infty} \Pr\left[X_{m+t} = \iota'\left(k'\right), m = \tau_{\beta,t} - t\right] = \Pr\left[\widetilde{X}_t = \iota'\left(k'\right)\right]$$

Therefore,

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

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

By Theorem 5.34, the state space for $\widetilde{X}_*$ is $\overline{\beta}$. If M is a minimal, then $\operatorname{im} \iota = \mathcal{S}$. Since
$\beta \subset \mathcal{S}$ and ι is 1-1, $\operatorname{im} \iota' \iota(\overline{s} = \overline{\beta}$. Thus, $\widehat{M}$ is minimal. $\square$

Theorem 5.35 allows us to easily show that the reduce construction on matrices of
section 5.2 is "natural", as well.

Theorem 5.36. *If $M \in \operatorname{Mat}_n(\mathbb{R})$ 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 $\tilde{s} = \overline{s_1}^{-1}(s_2)$, and $\left(\widehat{M}, p, i\right)$ is the reduction of M with respect to s, then
$M_2 = \widehat{M}$ and $p = p_2 p_1$. In addition, $i = i_1 i_2$ on $\ker\left(\widehat{M} - I\right)$.*

Proof. First, notice that, by Theorem 5.10, $\tilde{s}$ is open with respect to M_1, so that the state-
ment of the Theorem makes sense. If ι is the identity on S_n, given any n-dimensional
distribution, v, we may define a chain, X_*, so that M is an ι-consistent transition matrix
for X_* with initial distribution, v, i.e., $\Pr[X_0 = j] = v_j, \forall j \in S_n$.

If we first take $v = \frac{1}{n} J'$, so that all components are non-zero, the state space of X_* is
obviously all of S_n and M is minimal. By Theorem 5.35, taking $\beta_1 = \iota\left(s_1\right) = s_1$, M_1
is a minimal, ι_1-consistent transition matrix for $\overline{X}_* = \pi_{\beta_1}\left(X_*\right)$ and $\iota_1 = \iota\overline{s_1} = \overline{s_1}$ with
state space $\overline{\beta_1} = \overline{s_1}$. Likewise, taking $\beta_2 = \iota_1\left(\tilde{s}\right) = \overline{s_1}\left(\tilde{s}\right) = \overline{s_1} \cap s_2$, M_2 is a minimal, ι_2-
consistent transition matrix for $\widetilde{X}_* = \pi_{\beta_2}\left(\overline{X}_*\right)$ and $\iota_2 = \iota_1\overline{\tilde{s}}$ with state space $\overline{\beta_2} = \overline{\beta_1} \backslash \beta_2 =$
$\overline{s_1} \setminus \overline{s_1} \cap s_2 = \overline{s_1} \cap \overline{s_2} = \overline{s}$. Moreover, since $\operatorname{im} \iota_2 = \overline{\beta_2} = \overline{s}$, $\overline{s_1}\left(\overline{\tilde{s}}\right) = \iota_1\overline{\tilde{s}} = \iota_2 = \overline{s}$ as
enumerations, as well.

Taking $\beta = \beta_1 \cup \beta_2 = s_1 \cup (\overline{s_1} \cap s_2) = s_1 \cup s_2 = s$, $\widehat{M}$ is a minimal ι'-consistent transition matrix for $\widehat{X}_* = \pi_\beta(X_*)$ and $\iota' = \iota\overline{s} = \overline{s} = \iota_2$ with state space, $\overline{\beta} = \overline{s}$. By Theorem 5.33, $\widetilde{X}_* = \pi_{\beta_1}\left(\pi_{\beta_2}(X_*)\right) = \pi_\beta(X_*) = \widehat{X}_*$ with probability 1. Since M_2 and $\widehat{M}$ are both minimal transition matrices for $\widetilde{X}_* = \widehat{X}_*$, by Lemma 5.24, $M_2 = (P^\rho)^t\,\widehat{M}P^\rho$ for $\rho = \iota'^{-1}\iota_2$. Since $\iota' = \iota_2$, ρ is the identity permutation, $P^\rho = I$, and $M_2 = \widehat{M}$.

Returning to examine the p_i's, apply the same constructions as above, but now do not restrict v to any particular value. Theorem 5.35 also implies that $\Pr\left[\overline{X}_0 = \iota_1(j)\right] = (p_1 v)_j$ for all j. That is, the distribution vector for $\overline{X}_0$ is $p_1 v$. This implies that $\Pr\left[\widetilde{X}_0 = \iota_2(j)\right] = (p_2(p_1 v))_j$ for all j, i.e., the distribution vector for $\widetilde{X}_0$ is $p_2 p_1 v$. Likewise, the distribution vector for $\widehat{X}_0$ is pv. Since $\widetilde{X}_0 = \widehat{X}_0$, $p_2 p_1 v = pv$. Since this is true for all v, we must have $p_2 p_1 = p$.

Since $\overline{s_1}\left(\overline{\widetilde{s}}\right) = \overline{s}$ as enumerations, $\overline{s_1}\left(\overline{\widetilde{s}}_j\right) = \overline{s}_j$ for all j. Thus, $\iota_{\overline{s_1}}\iota_{\overline{s}}e_j = \iota_{\overline{s_1}}e_{\overline{\widetilde{s}}_j} = e_{\overline{s_1}(\overline{\widetilde{s}}_j)} = e_{\overline{s}_j} = \iota_{\overline{s}}e_j$ for all j, and $\iota_{\overline{s}} = \iota_{\overline{s_1}}\iota_{\overline{\widetilde{s}}}$. Therefore, $\pi_{\overline{s}}^t = \iota_{\overline{s}} = \iota_{\overline{s_1}}\iota_{\overline{\widetilde{s}}} = \pi_{\overline{s_1}}^t\pi_{\overline{\widetilde{s}}}^t = \left(\pi_{\overline{\widetilde{s}}}\pi_{\overline{s_1}}\right)^t$, and $\pi_{\overline{s}} = \pi_{\overline{\widetilde{s}}}\pi_{\overline{s_1}}$.

Now recall that Theorem 5.12 says that i, i_1, and i_2 are 1-1 correspondences between their corresponding sets of stable distributions, with left-inverses, $\pi_{\overline{s}}$, $\pi_{\overline{s_1}}$, and $\pi_{\overline{\widetilde{s}}}$, respectively. Thus, for any $v \in \ker\left(\widehat{M} - I\right)$, $i(v) \in \ker(M - I)$, so that there exist $v' \in \ker(M_1 - I)$ and $v'' \in \ker(M_2 - I) = \ker\left(\widehat{M} - I\right)$ with $i(v) = i_1(v')$, $v' = i_2(v'')$, and $i(v) = i_1(i_2(v''))$. Therefore, $v'' = \pi_{\overline{\widetilde{s}}}i_2 v'' = \pi_{\overline{\widetilde{s}}}\pi_{\overline{s_1}}i_1 i_2 v'' = \pi_{\overline{s}}i_1 i_2 v'' = \pi_{\overline{s}}i(v) = v$, so that $i(v) = i_1(i_2(v))$. Since v was an arbitrary stable vector in $\ker\left(\widehat{M} - I\right)$, the final part of the theorem is proven. $\square$

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 we 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.

First, we will establish some geometric preliminaries regarding directed spanning trees. In particular, we will show that 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, based on enumerating the weights of the directed spanning trees in the graph of M, exploiting the well-known Markov Chain Tree Theorem (Friedlin and Wentzell, 1984). When M is unichain (i.e., its associated graph has a single closed class), this 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.

6.1   Directed Spanning Trees

As given by Theorem 6.18, 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 a 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. *A directed graph, $G = (V, E, s, t)$, is a directed tree rooted at v if and only if*

- *v has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge, and*

- *the undirected graph associated with G does not contain any cycles.*

Proof. Consider a directed tree, G, rooted at v. That is, G contains a unique directed walk from any vertex to v, so that we may define the function, $l_G : V \to \mathbb{N}$, such that $l_G(v)$ is the length of the unique walk in G from u to v, with $l_G(v) = 0$. For any edge $(u, w) \in E$, there exists a walk from w to v, so the unique walk from u to v must be the concatenation of the edge (u, w) to the walk from w to v. If not, this walk is not unique. Thus, for any edge $(u, w) \in V$, $l_G(u) = l_G(w) + 1$.

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_0, 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.

Conversely, let G be a graph containing no undirected cycles, in which one vertex, v, has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge. Notice that the strongly connected components of G are all singleton sets, since, by definition, any pair of vertices in a strongly connected set contains a directed cycle containing them both. Since every vertex but v possesses an outgoing edge, G is unichain with unique closed class, $\{v\}$. By Lemma 1.1, from any vertex $u \in V \setminus \{v\}$, G contains a walk from u to v. If there were more than one such walk, the first vertex at which the walks diverged would have two outgoing edges, which is impossible. Thus, the walk is unique and G is a directed tree rooted at v. $\square$

We can therefore define the *parent* of any non-root vertex, $u \in V \setminus \{v\}$, by $t(\alpha)$, where $s(\alpha) = u$.

This leads to the following well-known result.

Corollary 6.2. *If $G = (V, E, s, t)$ is a directed tree, then $|E| = |V| - 1$.*

Proof. G must be rooted as some vertex, v, and by Theorem 6.1, there is a well-defined mapping, $\sigma : V \setminus \{v\} \to E$, taking each non-root vertex to its unique outgoing edge, so that $s(\sigma(w)) = w$, i.e., σ is a right-inverse of s. Given any edge, $\alpha \in E$, $s(\alpha) \in V \setminus \{v\}$, since v has no outgoing edge. In fact, $\sigma(s(\alpha)) = \alpha$, since each non-root vertex has a *unique* outgoing edge. Thus, σ is both a right- and left-inverse, hence, both 1-1 and onto, i.e., a 1-1 correspondence. In particular, $|E| = |V \setminus \{v\}| = |V| - 1$ $\square$

As with undirected graphs, a *directed subtree* is a subgraph which is also a directed tree, and a *directed spanning subtree* is a directed subtree which spans G (i.e., contains all vertices of G). We know from Lemma 1.1 that every directed graph contains at least one closed class. In this section, our focus will be to show that if a directed graph contains *exactly* one closed class, this is, if it is unichain, then it contains directed spanning 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 6.3. *For any directed graph $G = (V, E)$, G_T contains a star at v iff G is unichain and v is a vertex in its closed 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. By Lemma 1.1, G must contain at least one closed class, call it $\mathcal{C}$. Now observe that 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. In addition, there cannot be more than one closed class, since, by the same argument, v would have to be in all of them. $\square$

Lemma 6.3 says that if G is unichain, then it contains a directed walk from any vertex in G to each vertex in its closed 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 6.4. *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 a given root node, v. Each vertex reached by the algorithm is first discovered, then placed in a queue, eventually to be dequeued and processed. To process a vertex, w, the algorithm discovers and enqueues all undiscovered vertices adjacent to it. 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.

Algorithm 1 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 spanning subgraph, $G' = (V, E')$, is also directed tree rooted at v.

First, observe that any non-root vertex, $u \neq v \in V$, is eventually discovered (turned GRAY) by Algorithm 1. Since G_T contains a star at v, there is a path from u to v in G. Enumerating the vertices on this path $p_1, \ldots, p_l$ with $p_1 = v$, $p_l = u$, and $(p_i, p_{i-1}) \in E$, we may proceed, by induction, to show that every vertex on the path, including $p_l = u$, will be discovered. The initialization step of the algorithm guarantees that $p_1 = v$ is discovered, anchoring the induction. Now suppose p_{i-1} is discovered. Then p_{i-1} will be enqueued, guaranteeing that it will eventually be processed. When it is processed, p_i is examined, since $(p_i, p_{i-1}) \in E$, and either it is already GRAY, or it is turned GRAY at that point. In either case, we see that p_i will be discovered, completing the inductive step of the argument.

We now show that there is a path in G' from every vertex to v. Enumerating the non-root vertices $w_1, \ldots, w_m$ in the order they are discovered, we proceed by complete induction to show that there is a path from w_i to v in G' for all i. The first, w_1, is discovered because there is an edge, (w_1, v) in G. Since this edge is added to E', this gives a path from w_1 to v

---

Algorithm 1 BFS Tree

Q = new Queue()
$E' = \emptyset$
foreach $\{u \in V\}$
    color[u] = WHITE
color[v] = GRAY
Q.enqueue(v)
while (!Q.isEmpty()) {
    $u = Q$.dequeue()
    foreach $\{w \mid (w, u) \in E\}$
        if (color[w] = WHITE) {
            color[w] = GRAY
            E'.insert((w, u))
            Q.enqueue(w)
        }
    color[u] = BLACK
}
$G' = (V, E')$

---

in G'. Now suppose there is a path from w_k to v in G' for all $k < i$. When w_i is discovered, an edge (w_i, w_j) is added to E' from w_i to the vertex, w_j, currently being processed. Since all vertices are discovered before they are processed, w_j must have been discovered before w_i, so that $j < i$. By induction hypothesis, there is a path from w_j to v in G', and since $(w_i, w_j) \in E'$, there is a path from w_i to v in G', completing the induction step.

It remains to observe that the path in G' from each vertex to v is unique. The argument is similar to that in the proof of Theorem 6.1. Since an outgoing edge is added to G' only as it is discovered in the **while** loop of Algorithm 1, and each edge is discovered only once, the out-degree of each non-root vertex is 1. For any non-root vertex, u, if there were two distinct walks from u to v, the first vertex at which the walks differed would have at least two distinct outgoing edges in G', which is impossible. Thus, there is a *unique* directed walk in G' from every vertex to v, that is, G' is s directed tree. Since it is a subgraph of G on the same vertex set, G' is a spanning subtree of G rooted at v. □

Lemma 6.3 says that G contains a star at v iff G is unichain and v is a vertex in its closed class. Lemma 6.4 says that G_T contains a star at v iff G contains a directed spanning subtree rooted at v. Therefore,

Theorem 6.5. *A directed graph G contains a directed spanning subtree rooted at a vertex v iff G is unichain and v is in its closed 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 suggestively refer to as "1-regular" mappings. Specifically, for a 1-regular graph $G = (V, E)$, define $\mathrm{map}(G) = \sigma$ 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, $\mathrm{mat}(\sigma) \in \mathcal{M}$, so that $(\mathrm{mat}(\sigma))_{i,j} = 1$ iff $\sigma(j) = i$. Observe that each column j of $\mathrm{mat}(\sigma)$ is the standard basis vector $e_{\sigma(j)}$, so that $\mathrm{mat}(\sigma) = \left(e_{\sigma(1)} \ldots e_{\sigma(n)}\right)$. Conversely, for any matrix, $M \in \mathcal{M}$, we can define $\mathrm{map}(M) \in T$ such that $\left(\mathrm{map}(M)\right)(j) = i$ 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, M_i, with their corresponding graphs, are shown in Figure 6.1.

For $i \in S_n$, define

$$\mathcal{M}_i = \left\{ M \in \mathcal{M} \mid M_{j,j} = 1 \text{ iff } j = i \right\} \text{ and } T_i = \left\{ \mathrm{map}(M) \mid M \in \mathcal{M}_i \right\}$$

These sets correspond to 1-regular graphs with exactly one self-loop at v_i, such as M_3 and M_4 in Figure 6.1. For example, $M_3 \in \mathcal{M}_3$ and $M_4 \in \mathcal{M}_1$. Equivalently, $\mathrm{map}(M_3) \in T_3$)

Figure 6.1: Four 1-Regular Markov Matrices and Graphs

$$M_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad M_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
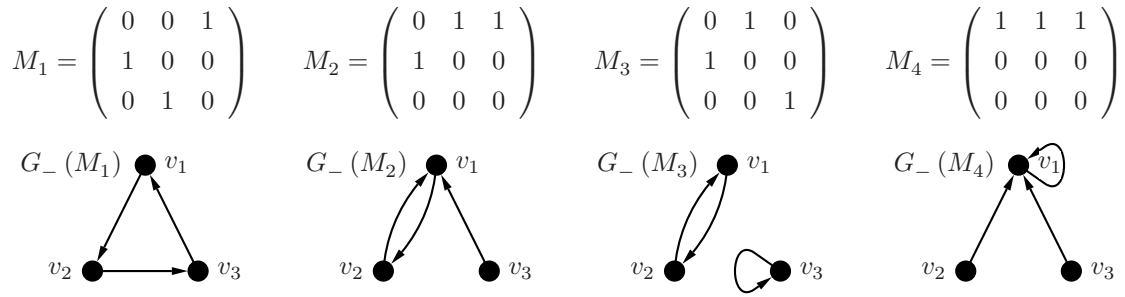
and $\mathrm{map}(M_4) \in T_1$.

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 $\overline{T}_i = \{\mathrm{map}(M) \mid M \in \overline{\mathcal{M}}_i\}$. 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.

Figure 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 $\overline{\mathcal{M}}_1$, or equivalently, $\mathrm{map}(M_i) \in \overline{T}_1$ for $i = 1, \dots, 3$. On the other hand, M_4 has two closed classes, $\{v_1\}$ and $\{v_2, v_3\}$, so it is not a member of $\overline{\mathcal{M}}_1$ and $\mathrm{map}(M_4) \notin \overline{T}_1$.

Figure 6.2: The set $\mathcal{M}_1$ of 3×3 matrices.

$$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad M_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$
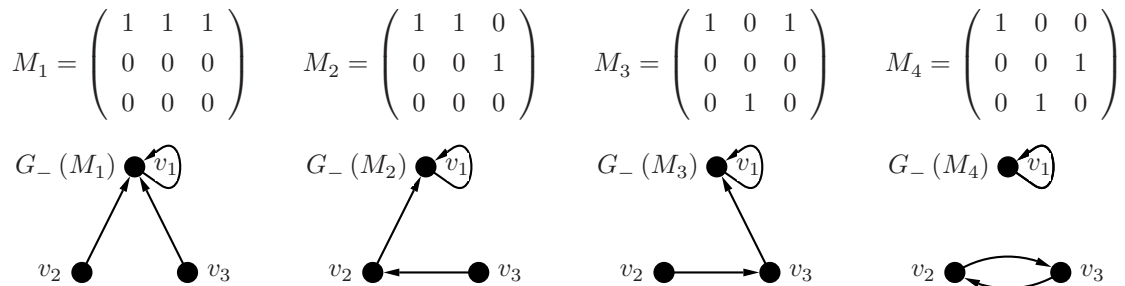
Now for $\sigma \in \overline{T}_i$, let $G_-^0(\sigma)$ be the graph obtained by removing the self-loop at v_i from

$G_-(\sigma)$. We will call the set of all such graphs $D_i = \{G^0_-(\sigma) \mid \sigma \in \overline{T}_i\}$. We thus have a 1-1 correspondence between maps $\sigma \in \overline{T}_i$, Markov matrices, $\mathrm{mat}(\sigma) \in \overline{\mathcal{M}}_i$, and directed graphs, $G^0_-(\sigma) \in D_i$. Figure 6.3, shows the Markov matrices, $\mathrm{mat}(\sigma)$ and graphs, $G^0_-(\sigma)$, associated with $\sigma \in \overline{T}_1$. Notice that all graphs are directed spanning trees rooted at v_1

Figure 6.3: The set D_1 of directed spanning trees with vertices in S_3, rooted at v_1

$$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$
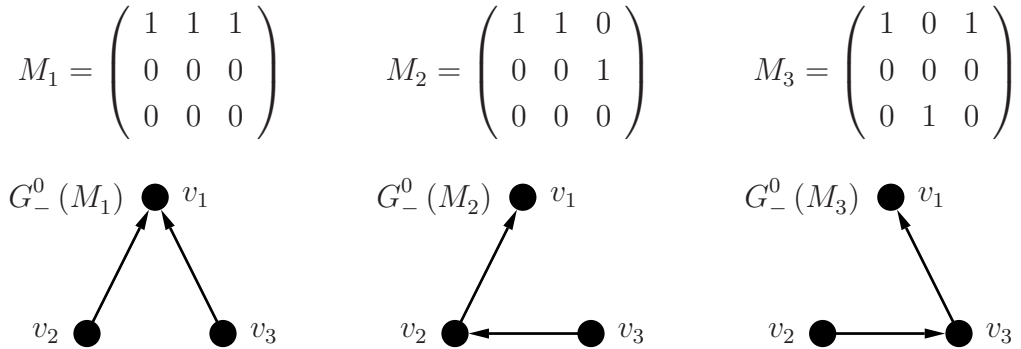
We now show that this correspondence gives us a way of enumerating all directed spanning trees rooted at v_i.

Theorem 6.6. *D_i is the set of all directed spanning trees on n vertices rooted at v_i.*

Proof. Let $\mathcal{D}_i$ denote the set of all directed spanning trees on n vertices rooted at v_i. Then we may view G^0_- as a mapping from $\overline{T}_i$ to D_i, which by definition is surjective. We wish to show that $D_i = \mathrm{im}\, G^0_- = \mathcal{D}_i$.

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 6.5, $G_-(\sigma)$ contains a directed spanning subtree, G', rooted at v_i. By Corollary 6.2, this graph has $n - 1$ edges, since $G_-(\sigma)$, and hence G', has n vertices. By definition, $G_-(\sigma)$ has n edges, so that, by removing the self-loop at v_i, we obtain a spanning subgraph, $G^0_-(\sigma)$, on n vertices and $n - 1$ edges.

To summarize, since $G_-(\sigma)$, G', and $G^0_-(\sigma)$ have the same vertex set. In addition, if we denote the corresponding edge sets by E, E', and E^0, respectively, we have that $E', E^0 \subset E \setminus \{(v_i, v_i)\}$, since G' is rooted at v_i and by construction of $G^0_-(\sigma)$. Since $|E'| = n - 1 = |E^0|$ and $|E \setminus \{(v_i, v_i)\}| = |E| - 1 = n - 1$, we must have $E' = E = E^0$, so that $G' = G^0_-(\sigma)$. That is, $D_i \subset \mathcal{D}_i$.

Moreover, every tree in $\mathcal{D}_i$ can be constructed in this fashion. Given a directed spanning tree, $G' \in \mathcal{D}_i$, with $V = \{v_1, \ldots, v_n\}$, rooted at v_i, add a self-loop at v_i to obtain a 1-regular graph, G, with associated $\sigma = \mathrm{map}(G)$. Since G' has no self-loops, G has exactly one self-loop at v_i, and $\sigma \in T_i$. Since the directed spanning tree, G' is a subgraph of G, by Theorem 6.5, 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, $\mathcal{D}_i \subset D_i$, so that $D_i = \mathcal{D}_i$. That is, D_i is the set of all directed spanning trees rooted at v_i. $\square$

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 6.7. 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 Figure 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 $w_M = \begin{pmatrix} \frac{3}{8} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$. $\square$

Lemma 6.8. *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 $(v_j, v_{\sigma(j)})$ for $j \in S_n \setminus \{i\}$. An edge, $(v_j, v_{\sigma(j)})$, 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 $0 \neq \Pi_{j \neq i} M_{\sigma(j),j} = W(M, \sigma)$. $\square$

By Lemma 6.8, 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 $G_-(M)$. That is,

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

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 6.9. *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 6.5 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 the 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 non-zero term in the sum

$\sum_{\sigma \in T(M,i)} W(M, \sigma)$. Since this is a sum of positive terms, $(w_M)_i \neq 0$. On the other hand, 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 $(w_M)_i = 0$.

Therefore, if M is unichain with v_i in the closed class of $G_-(M)$, then $(w_M)_i \neq 0$, and $w_M \neq 0$. We prove the converse implication by considering its contrapositive. That is, we show that, if M is not unichain, $(w_M)_i = 0$ for all i. If M is not unichain, by Theorem 6.5 $G(M)$ has no directed spanning subtree rooted at any vertex. So $T(M, i)$ is empty for all i, and $w_M = 0$. $\square$

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

formulas, given by Equations 6.5 and 6.6. The $(i, j)^{\text{th}}$ *minor* of N, $N^{i,j}$, is the $(n-1) \times (n-1)$ matrix obtained by removing the ith 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,

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

$$= \sum_{k=1}^{n} N_{k,j} C_N^{k,j} \tag{6.6}$$

Equation 6.5 is refered to as a Laplace expansion along row i, while Equation 6.6 is a Laplace expansion along column j.

Example 6.10. 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: $|M| = M_{1,1} C_M^{1,1} +$

$M_{2,1} C_M^{2,1} + M_{3,1} C_M^{3,1} = 0 \times C_M^{1,1} + 1 \times (-1)^{2+1} \begin{vmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} + 0 \times C_M^{3,1} = -1(\frac{1}{2} \times \frac{1}{2} - \frac{1}{4} \times \frac{1}{2}) =$

$-\frac{1}{8}$. $\square$

Theorem 6.11. *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* $|M| = v_1 \wedge \cdots \wedge v_n$,

$$(Nv_1) \wedge \cdots \wedge (Nv_n) = |N| \, v_1 \wedge \cdots \wedge v_n \tag{6.7}$$

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,

$$(\mathrm{adj}\, N)_{i,j} = C_N^{j,i} = (-1)^{i+j} \left| N^{j,i} \right| \tag{6.8}$$

It then turns out that the Laplace expansion formulas, 6.5 and 6.6, are equivalent to the following matrix equation (Wicks, 1996):

$$\mathrm{adj}(N)\, N = |N|\, I = N\, \mathrm{adj}(N) \tag{6.9}$$

Example 6.12. Using the matrix, M, from Example 6.10,

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

Similarly,

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

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 $\mathrm{adj}(M)M = M\, \mathrm{adj}(M) = \begin{pmatrix} -\frac{1}{8} & 0 & 0 \\ 0 & -\frac{1}{8} & 0 \\ 0 & 0 & -\frac{1}{8} \end{pmatrix} = |M|\, I.$ $\square$

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, remembering our convention that $\Lambda = M - I$, $(\overline{w}_M)_i \equiv \mathrm{adj}(\Lambda)_{i,i} = C_\Lambda^{i,i} = \left| \Lambda^{i,i} \right|$. In this section, we will show that $\overline{w}_M$ is closely related to the vector w_M, defined in section 6.1.3. This

will lead to a formula for the stable distribution of a unichain Markov matrix in terms of its directed spanning subtrees.

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

Example 6.13. For M from Example 6.10, $\Lambda = M - I = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{4} \\ 1 & -1 & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$, $\mathrm{adj}(\Lambda) =$

$\begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, so $\overline{w}_M = \begin{pmatrix} \frac{3}{8} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$. Moreover, $R_1(\Lambda) = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & -1 & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$. Using the

Laplace expansion formula to compute, $|R_1(\Lambda)| = 1 \times \begin{vmatrix} -1 & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -1 \times -\frac{1}{2} - \frac{1}{4} \times \frac{1}{2} =$

$\frac{3}{8} = (\overline{w}_M)_1$. $\square$

We now prove the result suggested by Example 6.13.

Lemma 6.14. *For any Markov matrix, M, $(\overline{w}_M)_i = |R_i(\Lambda)|$.*

Proof. By the Laplace expansion formula for the ith column,

$$
\begin{aligned}
|R_i(\Lambda)| &= \sum_{k=1}^{n} R_i(\Lambda)_{k,i} C_{R_i(\Lambda)}^{k,i} = \sum_{k=1}^{n} (e_i)_k \, C_{R_i(\Lambda)}^{k,i} \\
&= \sum_{k=1}^{n} [i = k] C_{R_i(\Lambda)}^{k,i} = C_{R_i(\Lambda)}^{i,i} = (-1)^{2i} |R_i(\Lambda)^{i,i}| \\
&= (-1)^{2i} |\Lambda^{i,i}| = \mathrm{adj}(\Lambda)_{i,i} = (\overline{w}_M)_i.
\end{aligned}
$$

$\square$

Comparing Examples 6.7 and 6.13, it appears that $\overline{w}_M = w_M$. We will now work to show that this is true, in general, with a minor caveat. We will do this by gradually rewriting $\overline{w}_M$ via a series of lemmas, until we obtain w_M.

Lemma 6.15. *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, $\Lambda(\mathrm{mat}(\sigma)) = \mathrm{mat}(\sigma) - I$.*

---

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

Proof. By Lemma 6.14, remembering that $\Lambda \equiv M - I$, we may first write $(\overline{w}_M)_i$ in terms of the columns of $R_i(\Lambda)$,

$$
\begin{aligned}
(\overline{w}_M)_i &= |R_i(\Lambda)| \\
&= R_i(\Lambda) e_1 \wedge \cdots \wedge R_i(\Lambda) e_i \wedge \cdots \wedge R_i(\Lambda) e_n \\
&= \Lambda e_1 \wedge \cdots \wedge e_i \wedge \cdots \wedge \Lambda e_n \quad (6.10)
\end{aligned}
$$

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 $\ker J$). To do so, we begin as follows:

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

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

$$
\Lambda e_j = \sum_{i \neq j} M_{i,j} e_i + \left(-\sum_{i \neq j} M_{i,j}\right) e_j = \sum_{i \neq j} M_{i,j}(e_i - e_j) = \sum_{i \neq j} M_{i,j} \overline{e}_{i,j} \quad (6.11)
$$

Applying Equation 6.11 to Equation 6.10, we have, by the multi-linearity of the determinant,

$$
\begin{aligned}
(\overline{w}_M)_i &= \sum_{s_1 \neq 1} M_{s_1,1} \overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \sum_{s_n \neq n} M_{s_n,n} \overline{e}_{s_n,n} \\
&= \sum_{s_1 \neq 1} \cdots \sum_{s_n \neq n} (M_{s_1,1} \cdots M_{s_n,n}) (\overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{s_n,n}) \\
&= \sum_{s_1 \neq 1} \cdots \sum_{s_n \neq n} \left(\Pi_{j \neq i} M_{s_j,j}\right) (\overline{e}_{s_1,1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{s_n,n}) \quad (6.12)
\end{aligned}
$$

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 6.12 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 6.12 may be rewritten as

$$
\begin{aligned}
(\overline{w}_M)_i &= \sum_{\sigma \in T_i} \left(\Pi_{j \neq i} M_{\sigma(j),j}\right) (\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n}) \\
&= \sum_{\sigma \in T_i} W(M, \sigma) (\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n}) \quad (6.13)
\end{aligned}
$$

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

$$
\begin{aligned}
\overline{e}_{\sigma(1),1} \wedge \cdots \wedge e_i \wedge \cdots \wedge \overline{e}_{\sigma(n),n} &= \left| \begin{array}{ccccc} \overline{e}_{\sigma(1),1} & \cdots & e_i & \cdots & \overline{e}_{\sigma(n),n} \end{array} \right| \\
&= \left| R_i \left(\begin{array}{ccccc} \overline{e}_{\sigma(1),1} & \cdots & \overline{e}_{\sigma(i),i} & \cdots & \overline{e}_{\sigma(n),n} \end{array} \right) \right| \\
&= \left| R_i \left(\begin{array}{ccc} e_{\sigma(1)} - e_1 & \cdots & e_{\sigma(n)} - e_n \end{array} \right) \right| \\
&= \left| R_i \left(\left(\begin{array}{ccc} e_{\sigma(1)} & \cdots & e_{\sigma(n)} \end{array} \right) - I \right) \right| \\
&= |R_i(\Lambda(\mathrm{mat}(\sigma)))| \qquad (6.14)
\end{aligned}
$$

Combining Equations 6.13 and 6.14 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 6.15 may be simplified significantly, once we prove the following lemma.

Lemma 6.16. *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 index of any non-root vertex is greater than that of its parent. 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 $(1,1)$-entry. By Theorems 2.3 and 6.11 d), permuting the rows and columns does not affect the determinant. So by Theorem 6.11 f), the determinant is the product of these 1×1 diagonal blocks, and $|R_i(\Lambda(\text{mat}(\sigma)))| = (-1)^{n-1}$.

Now suppose that $\sigma \notin \overline{T}_i$, so that $M = \text{mat}(\sigma)$ is not unichain, and has at least two closed classes. Pick two such closed classes, s_1 and s_2. By Lemma 2.11, there exists a permutation matrix, P, 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 $D_3 = M_2 - I = \Lambda_2$. In addition, $P^t R_i(\Lambda) P$ is 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 6.11 d), $|R_i(\Lambda)| = |P^t R_i(\Lambda) P|$. This is either $|R_{i'}(*_0 - I)| |\Lambda_1| |\Lambda_2|$, $|*_0| |R_{i'}(\Lambda_1)| |\Lambda_2|$, or $|*_0| |\Lambda_1| |R_{i'}(\Lambda_2)|$, by Theorem f). 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 $|R_i(\Lambda)| = 0$. $\square$

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

Theorem 6.17. *For any Markov matrix, M, $\overline{w}_M = (-1)^{n-1} w_M$.*

Proof. Focusing attention on the ith 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 6.15,
$$(\overline{w}_M)_i = \sum_{\sigma \in T_i} W(M, \sigma) |R_i(\Lambda(\text{mat}(\sigma)))|.$$

By Lemma 6.16, $|R_i(\Lambda(\text{mat}(\sigma)))| = 0$, if $\sigma \notin \overline{T}_i$. Thus, this simplifies to

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

Moreover, when $\sigma \in \overline{T}_i$, Lemma 6.16 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. $\square$

We now come to the key result of the chapter.

Theorem 6.18 (Markov Chain Tree Theorem). *For any unichain Markov matrix, M,* $\mathrm{stab}\, M = \{v_M\}$, *where*

$$(v_M)_i = \frac{1}{K}\left[\sum_{\sigma \in T(M,i)} W(M, \sigma)\right] = \frac{1}{K}(w_M)_i \tag{6.15}$$

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 2.12, $\ker \Lambda \neq 0$, so that Λ is not invertible. By Theorem 6.11 c), $|\Lambda| = 0$, and by Equation 6.9, $0 = |\Lambda|\, I = \mathrm{adj}\,(\Lambda)\,\Lambda$. Therefore, $0 = (\mathrm{adj}\,(\Lambda)\,\Lambda)^t = \Lambda^t\,(\mathrm{adj}\,(\Lambda))^t$. In other words, all rows of $\mathrm{adj}(\Lambda)$ are in $\ker \Lambda^t$. Likewise, since $J\Lambda = JM - J = 0$, $J \in \ker \Lambda^t$.

By Theorem 5.14, since M has 1 closed class, $\dim \ker \Lambda = 1$, which, by Theorem A.1, equals $\dim \ker \Lambda^t$. Since all vectors in a 1-dimensional subspace are multiples of any chosen non-zero member, each row of $\mathrm{adj}(\Lambda)$ must be a multiple of J. In other words, all entries in any given row must be equal. Equivalently, all columns of $\mathrm{adj}(\Lambda)$ are all equal.

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. Appealing again to Equation 6.9, $0 = |\Lambda|\, I = \Lambda\,\mathrm{adj}\,(\Lambda)$, so that the columns of $\mathrm{adj}(\Lambda)$ are all in $\ker \Lambda$, that is, $\overline{w}_M \in \ker \Lambda$.

Therefore, by Theorem 6.17, $w_M \in \ker \Lambda$. To this point, we could still have $\mathrm{adj}(\Lambda) = 0$. However, since M is unichain, Theorem 6.9 guarantees that $w_M \neq 0$. In particular, $\{w_M\}$ 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. By Corollary 5.15, v_M is the unique stable distribution of M. $\square$

Example 6.19. Continuing with the matrix, M, from Example 6.10, we obtain a normal-

ization constant of $K = (w_M)_1 + (w_M)_2 + (w_M)_3 = \frac{11}{8}$, so $v_M = \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix}$. We may easily

verify that v_M is a stable distribution, since $M v_M = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix} = \begin{pmatrix} \frac{3}{11} \\ \frac{4}{11} \\ \frac{4}{11} \end{pmatrix} =$

v_M. $\square$

Notice that Theorem 6.18 suggests that $(v_M)_i$ may be viewed as the conditional prob-
ability 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 sub-
trees, 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 pro-
cedure 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 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 equivalence class of a PMM, M_ϵ, may be specified by two real-valued matrices (i.e., its resistance, $R\left(M_\epsilon\right)$, and cost, $C\left(M_\epsilon\right)$),

- 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 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 "$\exists \delta > 0$ s.t. $Q(\epsilon)$ for $\epsilon \in [0, \delta]$". Likewise, "$Q(\epsilon)$ for $\epsilon \succ 0$" will mean "$\exists \delta > 0$ 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

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

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 $f_1 f_2 \simeq g_1 g_2$. $\square$

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 $\{f \in C^+[0,*] \mid \exists r, c \geq 0, \ f \simeq c\epsilon^r\}$. 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:

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

    d) 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 $R(f) = r$. $\square$

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 d), 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

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

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

$$= \frac{1}{c}\left(C(f) - [R(f) = R(g)]C(g) \right) = 1$$

Therefore, by Theorem 7.3 d), $f - g \in \mathcal{C}^+$, $C(f - g) = c = C(f) - [R(f) = R(g)]C(g)$, 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 in the limit,

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

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

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

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. $\square$

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

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

Likewise, by induction hypothesis,

$$C(f) = C(\overline{f}+f_k) = [R(f)=R(\overline{f})]C(\overline{f}) + [R(f)=R(f_k)]C(f_k)$$
$$= \sum_{i\in S_{k-1}}[R(f)=R(\overline{f})][R(\overline{f}=R(f_i))]C(f_i) + [R(f)=R(f_k)]C(f_k)$$

Now consider any $i \in S_{k-1}$ and observe that $R(f_i) = \min_{j \in S_k} R(f_j) \implies R(f_i) = \min_{j \in S_{k-1}} R(f_j)$. Therefore, by the formula for R given above, $R(f) = R(f_i) \implies R(f_i) = R(\overline{f})$, and $R(f) = R(f_i) \implies R(f) = R(\overline{f})$. This implies that $R(f) = R(f_i) \iff R(f) = R(\overline{f})$ and $R(f) = R(f_i)$, so that $R(f) = R(f_i) \iff R(f) = R(\overline{f})$ and $R(\overline{f}) = R(f_i)$. In other words,

$$
\begin{aligned}
[R(f) = R(f_i)] &= [R(f) = R(\overline{f}) \text{ and } R(\overline{f}) = R(f_i)] \\
&= [R(f) = R(\overline{f})][R(\overline{f}) = R(f_i)]
\end{aligned}
$$

for $i \in S_{k-1}$. Therefore,

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

as desired.

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

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

Likewise,

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

□

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, $C\,(M_\epsilon)$, where $C\,(M_\epsilon)_{i,j} = C\left((M_\epsilon)_{ij}\right)$.[3]

We will say that two perturbed matrices, M_ϵ and M'_ϵ 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 c), $M_\epsilon \simeq M'_\epsilon$ 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.

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

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

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

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

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

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

---

[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.

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):

$$C\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j}$$

$$= C\left(\left(M_\epsilon \widetilde{M}_\epsilon\right)_{i,j}\right) = C\left(\sum_{k \in S_m} (M_\epsilon)_{i,k} \left(\widetilde{M}_\epsilon\right)_{k,j}\right)$$

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

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

Similarly, by Corollary 7.5 and Theorem 7.4 d)

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

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

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

□

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 $(M_\epsilon)_{j,i} \not\simeq 0$, 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 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.[6] Notice that, since $c \cdot 0^r = c > 0$ iff $r = 0$ (and $c > 0$), $G_0(M_\epsilon)$ may also be thought of as the "zero-resistance" subgraph

---

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

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

[6] Although this notation conflicts somewhat with that given for undirected graphs introduced in section 1.1, the meaning is clear from the context.

of $G\left(M_\epsilon\right)$, that is, the weighted directed subgraph of $G\left(M_\epsilon\right)$ containing all edges, $\left(v_i, v_j\right)$, 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 $\mathrm{stab}\left(M_\epsilon\right)$ denote the unique stable distribution of M_ϵ for $\epsilon \succ 0$ given by Corollary 5.15.[7] Using the notation of Theorem 6.18, $\mathrm{stab}\left(M_\epsilon\right) = v_{M_\epsilon}$ for $\epsilon \succ 0$. We will show that $\mathrm{stab}\left(M_\epsilon\right)$ may be defined at $\epsilon = 0$ so that its entries are all in $\mathcal{C}^+$. In particular, $\mathrm{ssd}\left(M_\epsilon\right) \equiv \lim_{\epsilon \to 0^+} \mathrm{stab}\left(M_\epsilon\right)$ exists. We call this limit the *stochastically stable distribution* of M_ϵ. The set of indices, i, for which $\mathrm{ssd}\left(M_\epsilon\right)_i > 0$, or equivalently, for which $\mathrm{stab}\left(M_\epsilon\right)_i = 0$, are called the *stochastically stable states* of M_ϵ, and we define $\mathrm{sss}\left(M_\epsilon\right) \equiv \left\{i \mid \mathrm{ssd}\left(M_\epsilon\right)_i > 0\right\}$.

To prove this, we will need to extend the notation of chapter 6. Since $G\left(M_\epsilon\right)$ 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 $W\left(M_\epsilon, \sigma\right) = \Pi_{j \neq i}\left(M_\epsilon\right)_{\sigma(j), j}$. By Corollary 7.5 b), $W\left(M_\epsilon, \sigma\right) \in \mathcal{C}^+$. Thus, we may also define the *resistance of σ in M_ϵ* as $R\left(M_\epsilon, \sigma\right) \equiv R\left(W\left(M_\epsilon, \sigma\right)\right)$. Similarly, we define the *cost of σ in M_ϵ* as $C\left(M_\epsilon, \sigma\right) \equiv C\left(W\left(M_\epsilon, \sigma\right)\right)$. By Corollary 7.5 b), the resistance and cost of the tree, σ, satisfy the following equations:

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

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

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

Notice that, if $G_-\left(\sigma\right)$ is not a subgraph of $G_-\left(M_\epsilon\right)$, it does not contain some edge, $\left(v_{\sigma(j)}, v_j\right)$, so that $\left(M_\epsilon\right)_{\sigma(j), j} \simeq 0$. In terms of the resistance, the resistance of such an edge is ∞, so that the sum, $R\left(M_\epsilon, \sigma\right) = \infty$. Conversely, if the sum is infinite, the resistance of some edge is infinite, implying that $G_-\left(\sigma\right)$ is not a subgraph of $G_-\left(M_\epsilon\right)$. Likewise, $C\left(M_\epsilon, \sigma\right) = 0$ iff $G_-\left(\sigma\right)$ is not a subgraph of $G_-\left(M_\epsilon\right)$. Thus, corresponding to Equation 6.3, we may enumerate the set of directed spanning subtrees of $G_-\left(M_\epsilon\right)$, rooted at i, by

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

Also, let $T\left(M_\epsilon\right) \equiv \bigcup_{i \in S_n} T\left(M_\epsilon, i\right)$.

---

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

We now prove the statement above regarding $\mathrm{stab}\,(M_\epsilon)$. 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*

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

then

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

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

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

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

Proof. Using the notation of Theorem 6.18, fix a perturbed Markov matrix, M_ϵ.

Proof of part a): Since $R\,(M_\epsilon, \sigma) = \infty$, for any $\sigma \in \overline{T}_i \setminus T\,(M_\epsilon, i)$, we 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 $T\,(M_\epsilon) = \bigcup_i T\,(M_\epsilon, i)$,

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

Proof of part b): Now abbreviate w_{M_ϵ} as w_ϵ and set $K_\epsilon \equiv J w_\epsilon$.

For $\epsilon \succ 0$, by Theorem 6.18, we have $\mathrm{stab}\,(M_\epsilon) = v_{M_\epsilon} = \frac{w_\epsilon}{K_\epsilon}$ for each $\epsilon \succ 0$. It remains to show that we may extend this at 0 to a vector in $\mathrm{Pert}(n, 1)$. We first show that $w_\epsilon \in \mathrm{Pert}(n, 1)$.

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

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

In particular, by Corollary 7.5 a), we have that

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

and

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

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 $\mathrm{stab}(M_\epsilon) = \frac{w_\epsilon}{K_\epsilon} \in \mathrm{Pert}(n, 1)$.

Proof of part c): Theorem 7.4 e) also says that

$$R(\mathrm{stab}(M_\epsilon)_i) = R\left(\left(\frac{w_\epsilon}{K_\epsilon}\right)_i\right) = R((w_\epsilon)_i) - R(K_\epsilon) = r_i - r.$$

Proof of part d): Likewise, 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

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

By Theorem 6.17, we may compute $(w_\epsilon)_1$ in two ways. Using the adjoint formula:

$$
(w_\epsilon)_1 = (-1)^{3-1} \overline{w_\epsilon M_\epsilon} = \det \begin{vmatrix} -\epsilon - \epsilon^2 & \epsilon^3 \\ \epsilon^2 & -\epsilon - \epsilon^3 \end{vmatrix} = \epsilon^2 + \epsilon^3 + \epsilon^4
$$

Alternatively, we may sum over $\sigma \in \overline{T}_1$. Remember from Figure 6.3 that there are three such mappings: $2 \mapsto 1, 3 \mapsto 1$; $2 \mapsto 1, 3 \mapsto 2$; and $2 \mapsto 3, 3 \mapsto 1$.

$$
\begin{aligned}
(w_\epsilon)_1 &= \sum_{\sigma \in \overline{T}_1} (M_\epsilon, \sigma) = \sum_{\sigma \in \overline{T}_1} \prod_{p \neq 1} M_{\sigma(p),p} \\
&= (M_\epsilon)_{12}(M_\epsilon)_{13} + (M_\epsilon)_{12}(M_\epsilon)_{23} + (M_\epsilon)_{32}(M_\epsilon)_{13} = \epsilon^2 + \epsilon^4 + \epsilon^3
\end{aligned}
$$

Thus, $r_1 = 2$, so that only the first mapping is in $\overline{T}(M_\epsilon, 1)$ and contributes to the sums in $C\left(\text{stab}\left(M_\epsilon\right)\right)_1$.

Likewise, $(w_\epsilon)_2 = \epsilon^2 + 2\epsilon^4$ and $(w_\epsilon)_3 = \epsilon^2 + 2\epsilon^3$, so that $r_2 = r_3 = 2$, and $r = 2$. The mappings in $\overline{T}(M_\epsilon)$ correspond to the ϵ^2 terms in w_ϵ, which in turn correspond to the directed spanning trees with minimum resistance. Specifically, $\overline{T}(M_\epsilon)$ is given by the three mappings: $2 \mapsto 1, 3 \mapsto 1$; $3 \mapsto 1, 1 \mapsto 2$; and $2 \mapsto 1, 1 \mapsto 3$. Thus, it is easy to see that $C\left(\text{stab}\left(M_\epsilon\right)\right)_i = \frac{1}{3}$ for all i. □

As we have seen in section 5.3, indices correspond to states in a Markov process. The collection of indices, $i \in S_n$, such that $R\left(\text{stab}\left(M_\epsilon\right)\right)_i = 0$. are thus called the *stochastically stable* states of M_ϵ. These are precisely the indices for which the stochastically stable distribution, $v_0\left(M_\epsilon\right)$, has a non-zero component. As we will see in chapter 8, resistance may sometimes be interpreted as a "potential energy". Thus, stochastically stable states are then "minimal energy" states, which correspond to "stable" equilibria in physical system.

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 ("stochastic" equivalence) and show that equivalent PMMs have equal SSDs,

- observe that asymptotically equal PMMs are stochastically 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 stochastically 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 stochastically 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* stochastically equivalent, *denoted by $M_\epsilon \sim M'_\epsilon$, iff* $\mathrm{stab}\,(M_\epsilon) \simeq \mathrm{stab}\,(M'_\epsilon)$.

For example, we can show that asymptotically equal PMMs are stochastically 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 $R\left(\mathrm{stab}\left(M_\epsilon\right)\right)_i = R\left(\mathrm{stab}\left(M'_\epsilon\right)\right)_i$ and $C\left(\mathrm{stab}\left(M_\epsilon\right)\right)_i = C\left(\mathrm{stab}\left(M'_\epsilon\right)\right)_i$ 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. $\square$

Notice that the resistance and cost vectors of the stable distribution, and hence the stochastically stable distribution, are invariant under stochastic equivalence.

Theorem 7.11. *Given* $M_\epsilon, M'_\epsilon \in \mathrm{PMM}(n)$, *if* $M_\epsilon \sim M'_\epsilon$, *then*

a) $R\left(\mathrm{stab}\left(M_\epsilon\right)\right) = R\left(\mathrm{stab}\left(M'_\epsilon\right)\right)$,

b) $C\left(\mathrm{stab}\left(M_\epsilon\right)\right) = C\left(\mathrm{stab}\left(M'_\epsilon\right)\right)$, *and*

c) $\mathrm{ssd}\left(M_\epsilon\right) = \mathrm{ssd}\left(M'_\epsilon\right)$,

Proof. Parts a) and b) follow by Theorem 7.3. Theorem 7.4 a) gives part c). $\square$

When a uniform scaling of a PMM, M_ϵ, by $f \in \mathcal{C}^+$ yields another PMM, M'_ϵ, they are stochastically equivalent. That is, we have an analog of Lemma 5.3 for PMMs.

Theorem 7.12. *Given* $f \in \mathcal{C}^+ \setminus 0$ *and* $M_\epsilon \in \mathrm{PMM}(n)$, *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* $M'_\epsilon \in \mathrm{PMM}(n)$ *and* $M_\epsilon \sim M'_\epsilon$.

Proof. The proof is similar to that of Lemma 5.3. The real work is in proving that $M'_\epsilon \in \mathrm{PMM}(n)$. 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), $\left(M'_\epsilon\right)_{j,j} = \frac{(M_\epsilon)_{j,j}+f-1}{f} \in \mathcal{C}^+$. Thus, $\left(M'_\epsilon\right)_{i,j} \in \mathcal{C}^+$

It is now easy to show that $M_\epsilon \sim M'_\epsilon$. Since $f \not\approx 0$, $f(\epsilon) > 0$, $\Lambda'_\epsilon = \frac{1}{f(\epsilon)}\Lambda_\epsilon$, and $\ker \Lambda'_\epsilon = \ker \Lambda_\epsilon$, so that $\mathrm{stab}\,(M_\epsilon) = \mathrm{stab}\,\left(M'_\epsilon\right)$ for $\epsilon \succ 0$. In particular, by Lemma 7.2 a), $\mathrm{stab}\,(M_\epsilon) \simeq \mathrm{stab}\,\left(M'_\epsilon\right)$, so that $M_\epsilon \sim M'_\epsilon$. $\square$

In particular, when computing an SSD, we can always assume that $R\left(M_\epsilon\right)_{j,j} = 0$, for all j.

Corollary 7.13. *Given $M_\epsilon \in \mathrm{PMM}(n)$, if we define $M'_\epsilon = \frac{1}{2}\left(M_\epsilon - I\right) + I$, then $M'_\epsilon \in \mathrm{PMM}(n)$, $M_\epsilon \sim M'_\epsilon$, 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 7.12 with $f(\epsilon) = 2 \not\approx 0$. 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 7.12, $M'_\epsilon = \frac{1}{f}\left(M_\epsilon - I\right) + I = \frac{1}{2}\left(M_\epsilon - I\right) + I = \frac{1}{2}\left(M_\epsilon + I\right) \in \mathrm{PMM}(n)$, 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 $j \in S_n$. $\square$

We also should generalize the notion of D-equivalence from section 5.1 to apply to PMMs.

Definition 7.14. *For $D_\epsilon \in \mathrm{Pert}(n,m)$, $M_\epsilon \in \mathrm{PMM}(n)$, and $M'_\epsilon \in \mathrm{PMM}(m)$, we will say that M'_ϵ is D_ϵ-equivalent to M_ϵ, 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 $D_0 \mathrm{ssd}\left(M'_\epsilon\right) \propto \mathrm{ssd}\left(M_\epsilon\right)$.*

Proof. Since $M_\epsilon \sim_{D_\epsilon} M'_\epsilon$, $\mathrm{stab}\,(M_\epsilon), D_\epsilon \mathrm{stab}\left(M'_\epsilon\right) \in \ker \Lambda_\epsilon$. However, since M_ϵ is unichain, for $\epsilon \succ 0$, $\dim \ker \Lambda_\epsilon = 1$, so that $\mathrm{stab}\,(M_\epsilon) \propto D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)$. In other words, $f(\epsilon) \mathrm{stab}\,(M_\epsilon) = D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)$ for some function, f. Since

$$f(\epsilon) = f(\epsilon)J \mathrm{stab}\,(M_\epsilon) = JD_\epsilon \mathrm{stab}\left(M'_\epsilon\right) = \left\|D_\epsilon \mathrm{stab}\left(M'_\epsilon\right)\right\|_1,$$

we have, by Theorem 7.6 d) that f is in $Pert(1,1)$, i.e., $f \in \mathcal{C}^+$. By assumption, $R\left(f(\epsilon)\right) = 0$, so $C(f)\,\mathrm{ssd}\left(M_\epsilon\right) = D_0\,\mathrm{ssd}\left(M'_\epsilon\right)$, by Theorem 7.4 a). In particular, $D_0\,\mathrm{ssd}\left(M'_\epsilon\right) \propto \mathrm{ssd}\left(M_\epsilon\right)$. $\square$

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 $M_\epsilon \in \mathrm{PMM}(n)$, such that all communicating classes of M_0 are singletons, define $s \equiv \min_{j \notin \mathcal{T}} \min_{i \neq j} R\left(M_\epsilon\right)_{i,j}$, where $\mathcal{T}$ is the set of transient states of M_0. Likewise, let $c \equiv 2\max_{j \notin \mathcal{T}} \sum_{s=R(M_\epsilon)_{i,j}} C\left(M_\epsilon\right)_{i,j}$. If M_0 possesses more than one closed class, then $0 < s < \infty$ and $c > 0$. In addition, defining $f(\epsilon) \equiv c\epsilon^s$ and*

$$
(i_\epsilon)_{i,j} \equiv \begin{cases} 0 & \text{if } i \neq j \\ f(\epsilon) & \text{if } i = j \text{ and } j \in \mathcal{T} \\ 1 & \text{otherwise} \end{cases}
$$

then $i_\epsilon \in \mathrm{Pert}(n)$ and, using the notation of chapter 5, $\overline{M}_\epsilon \equiv (M_\epsilon)_{i_\epsilon} \in \mathrm{PMM}(n)$ and $\overline{M}_\epsilon \approx_{i_\epsilon} M_\epsilon$, for $\epsilon \succ 0$. Moreover, f satisfies the assumptions of Lemma 7.12 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 $1 > (M_0)_{j,j}$. By Theorem 7.4 a), this is equivalent to

$$
\begin{aligned}
0 &< 1 - (M_0)_{j,j} = \lim_{\epsilon \to 0^+}\left(1 - (M_\epsilon)_{j,j}\right) = \lim_{\epsilon \to 0^+}\sum_{i \neq j}(M_\epsilon)_{i,j} \\
&= \sum_{i \neq j}\lim_{\epsilon \to 0^+}(M_\epsilon)_{i,j} = \sum_{i \neq j}[R(M_\epsilon)_{i,j} = 0]C(M_\epsilon)_{i,j}
\end{aligned}
$$

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 Theorem 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 $s < \infty$.

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,

$$c = 2 \max_{\substack{j \notin \mathcal{T} \\ s = R(M_\epsilon)_{i,j}}} \sum C\left(M_\epsilon\right)_{i,j} > 0.$$

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 $0 < (i_\epsilon)_{j,j}$ for $j \in S_n$. We now show that, for $j \in S_n$ and $\epsilon \succ 0$, $(i_\epsilon)_{j,j} \left(1 - (M_\epsilon)_{j,j}\right) \leq 1$, so that, for $\epsilon \succ 0$, Lemma 5.3 implies that $\overline{M}_\epsilon \equiv (M_\epsilon)_{i_\epsilon}$ is Markov (by continuity, $\overline{M}_0$ is Markov, as well) and $\overline{M}_\epsilon \approx_{i_\epsilon} M_\epsilon$. Since $(M_\epsilon)_{j,j} \geq 0$, $(i_\epsilon)_{j,j} \left(1 - (M_\epsilon)_{j,j}\right) \leq (i_\epsilon)_{j,j}$. But either $(i_\epsilon)_{j,j} = 1$ or, since $R(f) = s > 0$, Theorem 7.4 a) implies that $(i_\epsilon)_{j,j} = f(\epsilon) \leq 1$ 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 $\left(\overline{M}_\epsilon\right)_{j,j} = (M_\epsilon)_{j,j} \in \mathcal{C}^+$ 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 $\left(\overline{M}_\epsilon\right)_{j,j} = f(\epsilon)(M_\epsilon)_{j,j} + 1 - f(\epsilon)$, 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 7.12 with respect to $\overline{M}_\epsilon$. Starting with a), notice that, for $i \neq j$, $\left(\overline{M}_\epsilon\right)_{i,j}$ either equals $f(\epsilon)(M_\epsilon)_{i,j}$, if $j \in \mathcal{T}$, or $(M_\epsilon)_{i,j}$, 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 7.12.

Now observe that, by Theorem 7.4 d), if $j \in \mathcal{T}$, $\left(\overline{M}_\epsilon\right)_{j,j} + f(\epsilon) - 1 = f(\epsilon)(M_\epsilon)_{j,j} \in \mathcal{C}^+$ and $R(f) \leq R\left(f(\epsilon)(M_\epsilon)_{j,j}\right)$. Thus, assumptions b) and c) are satisfied, in this case. On the other hand, if $j \notin \mathcal{T}$, $\left(\overline{M}_\epsilon\right)_{j,j} + f(\epsilon) - 1 = (M_\epsilon)_{j,j} + f(\epsilon) - 1 = f(\epsilon) - \left[\sum_{i \neq j} (M_\epsilon)_{i,j}\right]$.

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 $R\left(\sum_{i \neq j}(M_\epsilon)_{i,j}\right) = \min_{i \neq j} R\left(M_\epsilon\right)_{i,j} = s = R(f)$. In this case, $C\left(\sum_{i \neq j}(M_\epsilon)_{i,j}\right) = \sum_{i \neq j}[s = R(M_\epsilon)_{i,j}]C\left(M_\epsilon\right)_{i,j} = s =$

$R(f) \leq \frac{c}{2} < c = C(f)$. Otherwise, $R(f) = s < R\left(M_\epsilon\right)_{i,j}$ for all $i \neq j$, so that $R\left(\sum_{i \neq j}\left(M_\epsilon\right)_{i,j}\right) = \min_{i \neq j} R\left(M_\epsilon\right)_{i,j} > R(f)$. In either case, the conditions of Theorem 7.4 c) are satisfied, so that $f(\epsilon) - \left[\sum_{i \neq j}\left(M_\epsilon\right)_{i,j}\right] \in \mathcal{C}^+$ with resistance, $R(f)$. Thus, we have verified assumptions b) and c) of Theorem 7.12.

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 $g(\epsilon)\operatorname{stab}\left(M_\epsilon\right) = i_\epsilon\operatorname{stab}\left(M'_\epsilon\right)$ for some $g \in \mathcal{C}^+$, specifically, $g(\epsilon) = Ji_\epsilon\operatorname{stab}\left(M'_\epsilon\right) = \left\|i_\epsilon\operatorname{stab}\left(M'_\epsilon\right)\right\|_1$. Thus, we must show that $R(g) = 0$.

In particular, since $\left(i_\epsilon\operatorname{stab}\left(M'_\epsilon\right)\right)_j = \left(\operatorname{stab}\left(M'_\epsilon\right)\right)_j$ for $j \notin \mathcal{T}$, it suffices to show that $\operatorname{sss}\left(M'_\epsilon\right) \cap \overline{\mathcal{T}} \neq \emptyset$. In this case,

$$\lim_{0^+} g(\epsilon) \geq \lim_{0^+}\left(i_\epsilon\operatorname{stab}\left(M'_\epsilon\right)\right)_j = \lim_{0^+}\left(\operatorname{stab}\left(M'_\epsilon\right)\right)_j = \left(\operatorname{ssd}\left(M'_\epsilon\right)\right)_j > 0$$

for $j \in \operatorname{sss}\left(M'_\epsilon\right) \cap \overline{\mathcal{T}}$, so that $R(g) = 0$.

To see this, let $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ be the closed classes of $G_-\left(M_0\right)$, so that $\overline{\mathcal{T}} = \bigcup_{j=1}^m \mathcal{C}_i$. Likewise, let $\{\mathcal{C}'_1, \ldots, \mathcal{C}'_{m'}\}$ be the closed classes of $G_-\left(M'_0\right)$. By Theorems 5.17 and Corollary 5.15, $\operatorname{sss}\left(M'_\epsilon\right) = \bigcup_{j' \in \mathcal{J}'} \mathcal{C}'_{j'}$ for some $\mathcal{J}' \subset \{1, \ldots, m'\}$. Now observe that $G_-\left(M_0\right) \subset G_-\left(M'_0\right)$, so that, by Theorem 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 $j' \in \mathcal{J}'$, $\emptyset \neq \mathcal{C}'_{j'} \cap \bigcup_{j=1}^m \mathcal{C}_i \subset \operatorname{sss}\left(M'_\epsilon\right) \cap \overline{\mathcal{T}}$ $\square$

By repeatedly applying Lemma 7.16, we may then guarantee that M_ϵ always possesses a non-trivial communicating class.

Corollary 7.17. *Given $M_\epsilon \in \operatorname{PMM}(n)$, if M_0 possesses more than one closed class, there exists a diagonal $i_\epsilon \in \operatorname{Pert}(n)$ and $M'_\epsilon \in \operatorname{PMM}(n)$ such that $M_\epsilon \sim_{i_\epsilon} M'_\epsilon$ and M'_0 possesses a non-trivial communicating class or M'_0 is unichain.*

Proof. 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, we may apply Lemma 7.16 to obtain an i_ϵ-equivalent perturbed Markov matrix, M'_ϵ, so that $G_-\left(M'_0\right) \supset G_-\left(M_0\right)$ is a strictly larger graph. We may repeat this construction until either M'_0 possesses a non-trivial communicating class or M'_0 possesses only one closed class. This is guaranteed to terminate by the time $G_-\left(M'_0\right)$ is complete, if not sooner. $\square$

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 reduce 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 a stochastically equivalent one for which the off-diagonal zero-resistance terms are actually constant;

- we may also find a stochastically equivalent one for which all edges within an SCC (excluding self-loops) are constant; and

- we may construct a stochastically equivalent one which, for each non-zero resistance path from v_i to v_j in $G(M_\epsilon)$, contains a edge from v_i to v_j with the same resistance.

We begin by highlighting an important consequence of Theorem 7.7, that the stochastic equivalence class of a PMM depends only on its entries contained in some minimal resistance spanning subtree.

Theorem 7.18. *Given $M_\epsilon, M_\epsilon' \in \mathrm{PMM}(n)$, if*

a) $\overline{T}(M_\epsilon, i) = \overline{T}(M_\epsilon', i),$

b) $R(M_\epsilon)_{\sigma(j),j} = R(M_\epsilon')_{\sigma(j),j}, \forall j \neq i, \sigma \in \overline{T}(M_\epsilon, i)$

for every $1 \leq i \leq n$, then $R(\mathrm{stab}(M_\epsilon)) = R(\mathrm{stab}(M_\epsilon'))$. If we also have

c) $C(M_\epsilon)_{\sigma(j),j} = C(M_\epsilon')_{\sigma(j),j}, \forall j \neq \sigma(j), \sigma \in \overline{T}(M_\epsilon, i), \forall 1 \leq i \leq n,$

then $M_\epsilon \sim M_\epsilon'$.

Proof. Consider M_ϵ and M_ϵ' satisfying conditions a) and b) and define

$$\widetilde{T}_i \equiv \overline{T}(M_\epsilon, i) = \overline{T}(M_\epsilon', i) \text{ and } \widetilde{T} \equiv \bigcup_{1 \leq i \leq n} \widetilde{T}_i.$$

Condition a) then implies that

$$R\left(M_\epsilon, \sigma\right) = \sum_{j \neq i} R\left(M_\epsilon\right)_{\sigma(j),j} = \sum_{j \neq i} R\left(M'_\epsilon\right)_{\sigma(j),j} = R\left(M'_\epsilon, \sigma\right)$$

on $\widetilde{T}$.

Now observe that, for any $\sigma \in \widetilde{T}_i \subset \widetilde{T}$, using the notation of Theorem 7.7, we have $r_i = R\left(M_\epsilon, \sigma\right) = R\left(M'_\epsilon, \sigma\right) = r'_i$. By Theorem 7.7 a), $r = \min_i r_i = \min_i r'_i = r'$, as well. Theorem 7.7 c) then gives $R\left(\operatorname{stab}\left(M_\epsilon\right)\right) = R\left(\operatorname{stab}\left(M'_\epsilon\right)\right)$.

We now show that $\overline{T}(M_\epsilon) = \left\{\sigma \in \widetilde{T} \mid R\left(M_\epsilon, \sigma\right) = r\right\}$. The reverse inclusion is immediate. If $\sigma' \in \left\{\sigma \in \widetilde{T} \mid R\left(M_\epsilon, \sigma\right) = r\right\}$, then $\sigma' \in \widetilde{T}_i \subset \overline{T}_i$ for some i and $R\left(M_\epsilon, \sigma'\right) = r$. In particular, $\sigma' \in \overline{T}(M_\epsilon)$. Conversely, if $\sigma' \in \overline{T}(M_\epsilon)$, $\sigma' \in \overline{T}_i$ for some i and $R\left(M_\epsilon, \sigma'\right) = r$. By Theorem 7.7 a), $r_i = \min_{\sigma \in \overline{T}_i} R\left(M_\epsilon, \sigma\right) \leq r$. Theorem 7.7 a) also says that $r = \min_i r_i \leq r_i$, so that $R\left(M_\epsilon, \sigma'\right) = r = r_i$ and $\sigma' \in \widetilde{T}_i \subset \widetilde{T}$. Therefore, $\sigma' \in \left\{\sigma \in \widetilde{T} \mid R\left(M_\epsilon, \sigma\right) = r\right\}$, which implies that

$$\overline{T}(M_\epsilon) = \left\{\sigma \in \widetilde{T} \mid R\left(M_\epsilon, \sigma\right) = r\right\} = \left\{\sigma \in \widetilde{T} \mid R\left(M'_\epsilon, \sigma\right) = r\right\} = \overline{T}(M'_\epsilon).$$

If condition c) holds, as well, $C\left(M_\epsilon, \sigma\right) = C\left(M'_\epsilon, \sigma\right)$ for all $\sigma \in \widetilde{T}$, and

$$C\left(\operatorname{stab}\left(M_\epsilon\right)\right)_i = \frac{\sum_{\sigma \in \widetilde{T}_i} C\left(M_\epsilon, \sigma\right)}{\sum_{\sigma \in \overline{T}(M_\epsilon)} C\left(M_\epsilon, \sigma\right)} = \frac{\sum_{\sigma \in \widetilde{T}_i} C\left(M'_\epsilon, \sigma\right)}{\sum_{\sigma \in \overline{T}(M'_\epsilon)} C\left(M'_\epsilon, \sigma\right)} = C\left(\operatorname{stab}\left(M'_\epsilon\right)\right)_i.$$

Thus, by Theorem 7.3 c), $\operatorname{stab}\left(M_\epsilon\right) \simeq \operatorname{stab}\left(M'_\epsilon\right)$. In particular, by Definition 7.9, $M_\epsilon \sim M'_\epsilon$. $\square$

Using Theorem 7.18, we may give an alternate proof of Corollary 7.13.

Proof. Consider $M'_\epsilon = \frac{1}{2}I + \frac{1}{2}M_\epsilon$. Now $R\left(M_\epsilon\right)_{i,j} = R\left(M'_\epsilon\right)_{i,j}$ and $C\left(M_\epsilon\right)_{i,j} = C\left(M'_\epsilon\right)_{i,j}$ for $i \neq j$, since they only differ in their diagonal entries. Since a spanning tree does not contain self-loops, $R\left(M'_\epsilon, \sigma\right) = R\left(M_\epsilon, \sigma\right)$ for all $\sigma \in \overline{T}_i$ and i. In particular, $r'_i = \min_{\sigma \in \overline{T}_i} R\left(M'_\epsilon, \sigma\right) = \min_{\sigma \in \overline{T}_i} R\left(M_\epsilon, \sigma\right) = r_i$ and $\overline{T}\left(M'_\epsilon, i\right) = \overline{T}\left(M_\epsilon, i\right)$, for all i. Thus, Theorem 7.18 implies that $M_\epsilon \sim M'_\epsilon$. $\square$

Corollary 7.19. *For any $M_\epsilon \in \operatorname{PMM}(n)$ with $R\left(M_\epsilon\right)_{i,i} = 0$ for all i, there is an $M'_\epsilon \in \operatorname{PMM}(n)$, such that $M_\epsilon \simeq M'_\epsilon$ and $\left(M'_\epsilon\right)_{i,j} = C\left(M'_\epsilon\right)_{i,j} \epsilon^{R\left(M'_\epsilon\right)_{i,j}}$, for $i \neq j$.*

Proof. Setting $R \equiv R\left(M_\epsilon\right)$ and $C \equiv C\left(M_\epsilon\right)$, define

$$\left(M_\epsilon'\right)_{i,j} = \begin{cases} 1 - \sum_{k \neq j} C_{k,j} \epsilon^{R_{k,j}} & \text{if } i = j \\ C_{i,j} \epsilon^{R_{i,j}} & \text{otherwise} \end{cases}$$

where, by convention, $0\epsilon^\infty \equiv 0$. For $i \neq j$, Theorem 7.3 implies that $\left(M_\epsilon'\right)_{i,j} \in \mathcal{C}^+$ with $R\left(M_\epsilon'\right)_{i,j} = R_{i,j}$ and $C\left(M_\epsilon'\right)_{i,j} = C_{i,j}$. In particular, by Theorem 7.4 a), $\left(M_0'\right)_{i,j} = \left(M_0\right)_{i,j}$ for $i \neq j$.

On the diagonal, using the fact that $JM_\epsilon' = J = JM$,

$$\left(M_0'\right)_{j,j} = 1 - \sum_{i \neq j} \left(M_0'\right)_{i,j} = 1 - \sum_{i \neq j} \left(M_0\right)_{i,j} = \left(M_0\right)_{j,j} = C\left(M_\epsilon\right)_{i,i} > 0,$$

since $R\left(M_\epsilon\right)_{i,i} = 0$. This insures that $\left(M_\epsilon'\right)_{i,i} > 0$ for $\epsilon \succ 0$, with $\lim_{\epsilon \to 0^+} \frac{\left(M_\epsilon'\right)_{i,i}}{C(M_\epsilon)_{i,i}} = 1$, so that $\left(M_\epsilon'\right)_{i,i} \in \mathcal{C}^+$. Thus, $M_\epsilon' \in \mathrm{PMM}(n)$. Moreover, since $\left(M_0'\right)_{j,j} = \left(M_0\right)_{j,j}$, we must have $R\left(M_\epsilon'\right)_{j,j} = R\left(M_\epsilon\right)_{j,j}$ and $C\left(M_\epsilon'\right)_{j,j} = C\left(M_\epsilon\right)_{j,j}$ for all j, as well. Therefore, by Theorem 7.3 c) $M_\epsilon \simeq M_\epsilon'$. $\square$

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, $C\left(M_\epsilon\right)$ and $R\left(M_\epsilon\right)$, and we may assume that $R\left(M_\epsilon\right)_{i,i} = 0$, for all i.

Example 7.20. We may briefly illustrate the constructions of Corollaries 7.13 and 7.19 as follows:

$$\begin{pmatrix} 0 & \cdots \\ 1-\epsilon & \\ \epsilon & \end{pmatrix} \rightarrow \begin{pmatrix} \frac{1}{2} & \cdots \\ \frac{1}{2}-\frac{\epsilon}{2} & \\ \frac{\epsilon}{2} & \end{pmatrix} \rightarrow \begin{pmatrix} \frac{1}{2}-\frac{\epsilon}{2} & \cdots \\ \frac{1}{2} & \\ \frac{\epsilon}{2} & \end{pmatrix}$$

$\square$

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 $G_-\left(M_\epsilon\right)$ is unichain for $\epsilon \succ 0$, then there is a unique[8] $M'_\epsilon \in \mathrm{PMM}(n)$, 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 $\left(M'_\epsilon\right)_{j,j} = 1 - \sum_{i \neq j}\left(M_\epsilon\right)_{i,j}$ for $1 \leq j \leq n$, as required. Letting $c_j \equiv 1 - \sum_{i \neq j}\left(M_0\right)_{i,j} > 0$, by continuity, $\sum_{i \neq j}\left(M_\epsilon\right)_{i,j} < 1 - \frac{c_j}{2}$, so that $\left(M'_\epsilon\right)_{j,j} > \frac{c_j}{2} > 0$ for $\epsilon \succ 0$. Therefore, $\left(M'_\epsilon\right)_{j,j} \in \mathcal{C}^+$ and $R\left(M'_\epsilon\right)_{j,j} = 0$. Likewise, since $\left(M_0\right)_{j,j} = 1 - \sum_{i \neq j}\left(M_0\right)_{i,j} > 0$, $R\left(M_\epsilon\right)_{j,j} = 0$ for all j. In particular, since $R\left(M'_\epsilon\right)_{i,j} = R\left(M_\epsilon\right)_{i,j}$ for $i \neq j$, $R\left(M'_\epsilon\right) = R\left(M_\epsilon\right)$. Since $G_-\left(M_\epsilon\right)$ and $G_-\left(M'_\epsilon\right)$ contain an edge $\left(v_i, v_j\right)$ iff $R\left(M_\epsilon\right)_{j,i} = R\left(M'_\epsilon\right)_{j,i} < \infty$, they have the same unweighted graphs. Since $G_-\left(M_\epsilon\right)$ is unichain for $\epsilon \succ 0$, $G_-\left(M'_\epsilon\right)$ is as well. In particular, $M'_\epsilon \in \mathrm{PMM}(n)$. $\square$

Given a PMM, M_ϵ, and a communicating class of M_0, s, the following Lemma shows that we may assume that $\left(M_\epsilon\right)_{s,s}$ is constant off the diagonal.

Lemma 7.22.

If $M_\epsilon \in \mathrm{PMM}(n)$ with $R\left(M_\epsilon\right)_{i,i} = 0$ for all i, and s is contained in a communicating class of M_0, there is an $M'_\epsilon \in \mathrm{PMM}(n)$ such that $M'_\epsilon \sim M_\epsilon$ and, for $i \neq j$,

$$\left(M'_\epsilon\right)_{i,j} = \begin{cases} \left(M_0\right)_{i,j} & \text{if } i, j \in s \\[2mm] \left(M_\epsilon\right)_{i,j} & \text{otherwise} \end{cases}$$

Proof. We proceed by induction on the number of pairs, $(p, q) \in s \times s$ such that $0 < R\left(M_\epsilon\right)_{p,q} < \infty$. In this case, we define $M'_\epsilon = M_\epsilon - \left(M_\epsilon\right)_{p,q}\left(e_p e_q^t - e_q e_q^t\right)$. Since $R\left(M_\epsilon\right)_{p,q} \neq 0$, $p \neq q$, M'_ϵ is just M_ϵ with the (p, q)-entry removed. Moreover, $G\left(M'_\epsilon\right)$ is $G\left(M_\epsilon\right)$ with edge, $\left(v_p, v_q\right)$, removed. However, since $p, q \in s$, $\left(v_p, v_q\right) \in \mathcal{P}\left(M'_0\right) \subset \mathcal{P}\left(M'_\epsilon\right)$, and $\mathcal{P}\left(M'_\epsilon\right) = \mathcal{P}\left(M_\epsilon\right)$. In particular, M'_ϵ is unichain. Since $\left(M'_\epsilon\right)_{p,q} = 0$ and $\left(M'_\epsilon\right)_{q,q} = \left(M_\epsilon\right)_{p,q} + \left(M_\epsilon\right)_{q,q} \in \mathcal{C}^+$, $M'_\epsilon \in \mathrm{PMM}(n)$.

By Theorem 7.18, to show that $M'_\epsilon \sim M_\epsilon$, it suffices to verify that edge $\left(v_p, v_q\right)$ is not part of any minimal resistance spanning tree of M_ϵ. Specifically, if $\sigma \in T\left(M_\epsilon, k\right)$ 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., $\sigma \notin \overline{T}\left(M_\epsilon, k\right)$). Therefore, assume we are given $\sigma \in T\left(M_\epsilon, k\right)$ with $\sigma(q) = p$ and consider the associated directed spanning tree, $\mathcal{T}$.

---

[8]Up to equality in $\mathrm{Mat}_n\left(\mathcal{C}\right)$.

Let $\{s^1, \ldots, s^m\}$ be the strongly connected components of $G(M_0)$. As in the proof of Lemma 6.16, we may re-index the matrix so that the index corresponding to any non-root vertex is greater than that of its parent in $\mathcal{T}$, i.e., so that $\sigma(j) < j$ for $j \neq k$, and $k = 1$. By re-indexing again, we may further assume that $\min s^q < \min s^r \Rightarrow q < r$. In particular, we must have $1 \in s^1$.

We may now proceed by induction to construct a new directed spanning tree, $\overline{\mathcal{T}}$ corresponding to a regular mapping, $\overline{\sigma}$, such that $R(M_\epsilon, \overline{\sigma}) < R(M_\epsilon, \sigma)$. Specifically, we will construct a spanning tree rooted at 1 which only contains 0-resistance edges within each s^q. Moreover, there will be at most one edge in $\overline{\mathcal{T}}$ between distinct communicating classes, i.e., $\forall q > 1, \exists j_q \in s^q$ s.t. $\overline{\sigma}(j_q) = \sigma(j_q) \in s^u$ and $u < q$. Intuitively, we will choose edges of $\mathcal{T}$ to build a directed tree on the communicating classes (specified by j_q), which will serve to link a set of 0-resistance spanning trees spanning each class into an directed tree rooted at 1.

Since s^1 is a strongly connected component of $G(M_0)$, by Theorem 6.5, it contains an directed tree rooted at v_1 consisting entirely of 0-resistance edges spanning $G(M_0)|_{s^1}$, which defines $\overline{\sigma}(q)$ for $q \in s^1$. For any $1 < r \leq m$, assume that we have defined $\overline{\sigma}$ over $\bigcup_{q=1}^{r-1} s^q$ and a sequence $j_q \in s^q$ such that $\forall 1 < q < r, \overline{\sigma}(j_q) = \sigma(j_q) \in s^u$ for some $u < q$. It suffices to show that we may extend $\overline{\sigma}$ to s^r and define $j_r \in s^r$ so that $\overline{\sigma}(j_r) = \sigma(j_r) \in s^u$ for some $u < r$.

Let $j_r = \min s^r$. Since $\sigma(j_r) < r$, if $\sigma(j_r) \in s^u$, $\min s^u < j_r = \min s^r$, so that $u < r$. As before, Theorem 6.5 guarantees the existence of a directed tree rooted at j_r spanning $G(M_0)|_{s^r}$. This defines $\overline{\sigma}$ on $s^r - \{j_r\}$ and we take $\overline{\sigma}(j_r) = \sigma(j_r)$. By induction, we eventually obtain a subgraph of $G(M_\epsilon)$ consisting of $n - 1$ edges which contains a path from each vertex of $G(M_\epsilon)$ to 1, which is thus a 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, $R(\overline{\sigma}, M_\epsilon) \leq R(M_\epsilon, \sigma) - R(M_\epsilon)_{p,q} < R(M_\epsilon, \sigma)$. 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 $T(M_\epsilon, k) = T(M_\epsilon', k)$ 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, Theorem 7.18 implies that $M_\epsilon' \sim M_\epsilon$. By repeating this construction sufficiently many times, we obtain $M_\epsilon' \sim M_\epsilon$ such that $R(M_\epsilon')_{i,j} = 0$ or ∞ for $i \neq j \in s$. By applying Corollary 7.19, we obtain a new $M_\epsilon' \sim M_\epsilon$

such that $\left(M'_\epsilon\right)_{i,j} = C\left(M_\epsilon\right)_{i,j} = \left(M_0\right)_{i,j}$ for $i \neq j \in s$. $\square$

To illustrate, in Example 7.8, we may drop the $(3,2)$– and $(2,3)$–entries (adjusting the diagonal entries accordingly) without changing its stochastically stable distribution. We saw explicitly that these entries are not in any of the minimal resistance spanning trees.

7.6   Reduction of PMMs

We now wish to generalize the reduce 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, $G_-\left(M_0\right)$ of $G_-\left(M_\epsilon\right)$.

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 4.4 to PMMs, proving that $\left(I - \left(M_\epsilon\right)_{s,s}\right)^{-1} \in \mathrm{Pert}(|s|)$ 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 reduce 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 $\widehat{M}_\epsilon \in \mathrm{PMM}(|\overline{s}|)$;

- 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 stochastic equivalence of PMMs; 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 reduce 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 7.23. *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 4.4 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, $|\pi_s\,(I - M_0)\,\imath_s| \neq 0$, so that $|\pi_s\,(I - M_\epsilon)\,\imath_s| \neq 0$ for $\epsilon \succeq 0$. Thus, by the contrapositive of Lemma 4.4 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. $\square$

Lemma 7.23 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.

First, observe that for $M_\epsilon \in \mathrm{PMM}(n)$, $G_-\,(M_\epsilon)$ is constant for $\epsilon \succ 0$. Thus, for $s \subset S_n$, $\mathcal{P}_{M_\epsilon}(s, i, j, l) \subset \mathcal{S}_n\,(k)$ is a fixed subset, independent of ϵ. 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:

$$R\left(M_\epsilon, \sigma\right) = R\left(W\left(M_\epsilon, \sigma\right)\right)$$
$$= R\left(\Pi_{t\in S_k}\left(\left(M_\epsilon\right)_{\sigma_t,\sigma_{t-1}}\right)\right) = \sum_{t\in S_k} R\left(\left(M_\epsilon\right)_{\sigma_t,\sigma_{t-1}}\right) \tag{7.3}$$

$$C\left(M_\epsilon, \sigma\right) = C\left(W\left(M_\epsilon, \sigma\right)\right)$$
$$= C\left(\Pi_{t\in S_k}\left(\left(M_\epsilon\right)_{\sigma_t,\sigma_{t-1}}\right)\right) = \Pi_{t\in S_k} C\left(\left(M_\epsilon\right)_{\sigma_t,\sigma_{t-1}}\right) \tag{7.4}$$

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 7.24. *For any $M_\epsilon \in \mathrm{PMM}(n)$, $s \subset S_n$, and $i, j \in S_n$,*

$$r \equiv \min_{\sigma\in\mathcal{S}_n(s,i,j)} R\left(M_\epsilon, \sigma\right) = \min_{\sigma\in\mathcal{P}_{M_\epsilon}(s,i,j)} R\left(M_\epsilon, \sigma\right)$$

exists. In addition, $\mathcal{P}_{M_\epsilon}\left(s, i, j\right) \neq \emptyset$ iff $r < \infty$ iff

$$\overline{\mathcal{P}}_{M_\epsilon}\left(s, i, j\right) \equiv \left\{\sigma \in \mathcal{S}_n\left(s, i, j\right) \mid R\left(M_\epsilon, \sigma\right) = r\right\} \subset \mathcal{P}_{M_\epsilon}\left(s, i, j\right).$$

Proof. If $\mathcal{P}_{M_\epsilon}\left(s, i, j\right) = \emptyset$, then $R\left(M_\epsilon, \sigma\right) = \infty$ for all $\sigma \in \mathcal{S}_n\left(s, i, j\right)$. In this case, $\min_{\sigma\in\mathcal{S}_n(s,i,j)} R\left(M_\epsilon, \sigma\right) = \infty$ and we set $\overline{\mathcal{P}}_{M_\epsilon}\left(s, i, j\right) \equiv \emptyset$, consistent with our convention that $\min \emptyset = \infty$.

Otherwise, assume that $\mathcal{P}_{M_\epsilon}\left(s, i, j\right) \neq \emptyset$, and define $r_k = \min_{\sigma\in\mathcal{S}_n(s,i,j,k)} R\left(M_\epsilon, \sigma\right)$. This is a finite set, so the minimum is well-defined. Moreover, by the "pigeon hole" principle, any walk of length greater than n must visit at least one vertex more than once. Thus, such walks must contain a cycle, and hence have weight greater than or equal to that of a walk of length less than or equal to n. In particular, $r_k \leq r_n$ for all k. Thus,

$$r \equiv \min_{1\leq k\leq n} r_k = \min_k r_k$$
$$= \min_k \min_{\sigma\in\mathcal{S}_n(s,i,j,k)} R\left(M_\epsilon, \sigma\right) = \min_{\sigma\in\mathcal{S}_n(s,i,j)} R\left(M_\epsilon, \sigma\right)$$

Since $\mathcal{P}_{M_\epsilon}(s,i,j) \neq \emptyset$, $\emptyset \neq \mathcal{P}_{M_\epsilon}(s,i,j,k) \subset \mathcal{S}_n(s,i,j,k)$ for some k, so that $r_k < \infty$, and therefore, $r < \infty$.

We may then define $\overline{\mathcal{P}}_{M_\epsilon}(s,i,j) \equiv \{\sigma \in \mathcal{S}_n(s,i,j) \mid R(M_\epsilon, \sigma) = r\}$. For any $\sigma \in \mathcal{S}_n(s,i,j)$, $\sigma \in \mathcal{P}_{M_\epsilon}(s,i,j)$ iff $R(M_\epsilon, \sigma) < \infty$. In particular, if $\sigma \in \overline{\mathcal{P}}_{M_\epsilon}(s,i,j)$, $R(M_\epsilon, \sigma) = r < \infty$, so that $\sigma \in \mathcal{P}_{M_\epsilon}(s,i,j)$, i.e., $\overline{\mathcal{P}}_{M_\epsilon}(s,i,j) \subset \mathcal{P}_{M_\epsilon}(s,i,j)$. Notice this implies that $\min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,i,j)} R(M_\epsilon, \sigma) \leq r$. However, since $\mathcal{P}_{M_\epsilon}(s,i,j) \subset \mathcal{S}_n(s,i,j)$,

$$\min_{\sigma \in \mathcal{S}_n(s,i,j)} R(M_\epsilon, \sigma) \leq \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,i,j)} R(M_\epsilon, \sigma),$$

so that $\min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,i,j)} R(M_\epsilon, \sigma) = r$.

We have shown that $\mathcal{P}_{M_\epsilon}(s,i,j) \neq \emptyset \Rightarrow r < \infty \Rightarrow \overline{\mathcal{P}}_{M_\epsilon}(s,i,j) \subset \mathcal{P}_{M_\epsilon}(s,i,j)$. However, $\overline{\mathcal{P}}_{M_\epsilon}(s,i,j) \neq \emptyset$. Thus, $\overline{\mathcal{P}}_{M_\epsilon}(s,i,j) \subset \mathcal{P}_{M_\epsilon}(s,i,j) \Rightarrow \mathcal{P}_{M_\epsilon}(s,i,j) \neq \emptyset$, so that all three statements are equivalent. $\square$

In other words, $\overline{\mathcal{P}}_{M_\epsilon}(s,i,j)$ represents the set of minimum resistance walks in $G(M_\epsilon)$ from v_j to v_i passing only through V_s. We will partition this set of minimum resistance walks by their length, to define

$$\overline{\mathcal{P}}_{M_\epsilon}(s,i,j,k) = \mathcal{S}_n(s,i,j,k) \cap \overline{\mathcal{P}}_{M_\epsilon}(s,i,j)$$

as well.

We now show that $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathrm{Pert}$ and give formulas for its resistance and cost matrices in terms of minimum resistance walks.

Theorem 7.25. *For any $M_\epsilon \in \mathrm{PMM}(n)$ and $s \subset S_n$ an open set with respect to M_0, the partial sums, $(M_\epsilon)_{s,s}^{(u)} \equiv \sum_{k=0}^{u-1} (M_\epsilon)_{s,s}^k$, converge uniformly to $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathrm{Pert}$. Moreover,*

$$R\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)_{i,j} = \min_{\sigma \in \mathcal{S}_n\left(s,s_i,s_j\right)} R(M_\epsilon, \sigma) \text{ and}$$

$$C\left(\left(I - (M_\epsilon)_{s,s}\right)^{-1}\right)_{i,j} = \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}\left(s,s_i,s_j\right)} C(M_\epsilon, \sigma)$$

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 $(M_\epsilon)_{s,s}^{(u)} = \sum_{k=0}^{u-1} (M_\epsilon)_{s,s}^k$, are continuous and non-negative. By assumption, s is open with respect to M_0, and by Lemma 7.23 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 4.4, $(M_\epsilon)^{(u)}_{s,s}$ 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, by the proof of Theorem 4.4, parts a) and b), $M_\epsilon^{(u)}$ 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 $\left(I - (M_\epsilon)_{s,s}\right)^{-1} \in$ Pert, we must compute the resistance and cost of each of its entries. Consider the $(i, j)^{\text{th}}$ entry of $(M_\epsilon)^{(u)}_{s,s}$, for some fixed $i, j \in S_n$. By Lemma 4.3 b) and Corollary 7.5 a),

$$\left((M_\epsilon)^{(u)}_{s,s}\right)_{i,j} = e_i^t (M_\epsilon)^{(u)}_{s,s} e_j = \sum_{k=0}^{u-1} e_i^t (M_\epsilon)^k_{s,s} e_j = \sum_{k=0}^{u-1} \sum_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j,k)} W(M_\epsilon, \sigma) \in \mathcal{C}^+$$

Therefore, by Corollary 7.5 a) again,

$$\begin{aligned}
r_u &\equiv R\left((M_\epsilon)^{(u)}_{s,s}\right)_{i,j} = R\left(\sum_{k=0}^{u-1} \sum_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j,k)} W(M_\epsilon, \sigma)\right) \\
&= \min_{0 \leq k \leq u-1} \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j,k)} R(M_\epsilon, \sigma)
\end{aligned}$$

Also observe that

$$\begin{aligned}
\inf_u r_u &= \inf_u \min_{0 \leq k \leq u-1} \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j,k)} R(M_\epsilon, \sigma) \\
&= \inf_u \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j,u)} R(M_\epsilon, \sigma) \\
&= \inf_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j)} R(M_\epsilon, \sigma) = \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s,s_i,s_j)} R(M_\epsilon, \sigma)
\end{aligned}$$

Now take

$$u' = \min \left\{ u \mid \overline{\mathcal{P}}_{M_\epsilon}\left(s, s_i, s_j, u\right) \neq \emptyset \right\} \text{ and}$$
$$\sigma' \in \overline{\mathcal{P}}_{M_\epsilon}\left(s, s_i, s_j, u'\right)$$

By Theorem 7.25, we then have

$$r_{u'+1} = \min_{0 \leq k \leq u'} \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s, s_i, s_j, k)} R\left(M_\epsilon, \sigma\right) \leq R\left(M_\epsilon, \sigma'\right) = \min_{\sigma \in \mathcal{P}_{M_\epsilon}(s, s_i, s_j)} R\left(M_\epsilon, \sigma\right) = \inf_u r_u$$

Thus, $r_{u'+1} = \inf_u r_u$, so that $r \equiv \min_u r_u = r_{u'+1}$ is well-defined and $r \leq R\left(M_\epsilon, \sigma\right)$, for all $\sigma \in \mathcal{P}_{M_\epsilon}(s, s_i, s_j)$.

Now consider $\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j}$. By Theorem 7.4 e), this is in $\mathcal{C}^+$. For $u > u' + 1$, we have

$$\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j} = \epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u'+1)}\right)_{i,j} + \epsilon^{-r} \sum_{k=u'+1}^{u-1} \sum_{\sigma \in \mathcal{P}_{M_\epsilon}(s, s_i, s_j, k)} W\left(M_\epsilon, \sigma'\right).$$

so that, by Corollary 7.5 a) and Theorem 7.4 e),

$$R\left(\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j}\right) = R\left(\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u'+1)}\right)_{i,j}\right) = 0$$

Therefore,

$$c_u \equiv \lim_{\epsilon \to 0^+} \epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j}$$

$$= C\left(\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j}\right) \qquad \text{by Theorem 7.4 a)}$$

$$= C\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j} \qquad \text{by Theorem 7.4 e)}$$

$$= \sum_{0 \leq k \leq u-1} \sum_{\sigma \in \mathcal{P}_{M_\epsilon}(s, s_i, s_j, k)} [R\left(M_\epsilon, \sigma\right) = r] C\left(M_\epsilon, \sigma\right) \quad \text{by Corollary 7.5 a)}$$

$$= \sum_{0 \leq k \leq u-1} \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}(s, s_i, s_j, k)} C\left(M_\epsilon, \sigma\right) \qquad \text{by Lemma 7.24}$$

We may argue as before that $\epsilon^{-r}\left(\left(M_\epsilon\right)_{s,s}^{(u)}\right)_{i,j}$ is continuous on some interval, $[0, \delta]$. Moreover, for $u > u'+1$, this sequence converges uniformly to $\epsilon^{-r}\left(I - \left(M_\epsilon\right)_{s,s}\right)_{i,j}^{-1}$, which

is therefore also continuous on $[0, \delta]$. In particular, $\lim_{0^+} \epsilon^{-r} \left(I - (M_\epsilon)_{s,s} \right)^{-1}_{i,j}$ exists, and we may interchange limits to compute

$$
\begin{aligned}
\lim_{0^+} \epsilon^{-r} \left(I - (M_\epsilon)_{s,s} \right)^{-1}_{i,j} &= \lim_{0^+} \lim_{u \to \infty} \epsilon^{-r} \left((M_\epsilon)^{(u)}_{s,s} \right)_{i,j} \\
&= \lim_{u \to \infty} \lim_{0^+} \epsilon^{-r} \left((M_\epsilon)^{(u)}_{s,s} \right)_{i,j} \\
&= \lim_{u \to \infty} c_u = \lim_{u \to \infty} \sum_{0 \le k \le u-1} \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}(s,s_i,s_j,k)} C(M_\epsilon, \sigma)
\end{aligned}
$$

Since it is a sum of positive terms, $c_u \ge c_{u'+1} > 0$ for $u > u'$. In particular,

$$
\lim_{0^+} \epsilon^{-r} \left(I - (M_\epsilon)_{s,s} \right)^{-1}_{i,j} = \lim_{u \to \infty} c_u > 0.
$$

Thus, $R \left(\left(I - (M_\epsilon)_{s,s} \right)^{-1}_{i,j} \right) = r$, $\left(I - (M_\epsilon)_{s,s} \right)^{-1} \in \text{Pert}$, and

$$
\begin{aligned}
C \left(\left(I - (M_\epsilon)_{s,s} \right)^{-1} \right)_{i,j} &= \lim_{u \to \infty} \sum_{0 \le k \le u-1} \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}(s,s_i,s_j,k)} C(M_\epsilon, \sigma) \\
&= \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}(s,s_i,s_j)} C(M_\epsilon, \sigma).
\end{aligned}
$$

$\square$

Corollary 7.26. *If $M_\epsilon, M'_\epsilon \in \text{PMM}(n)$, $s \subset S_n$ is open with respect to M_0 and M_0', and $(M_\epsilon)_{s,s} \simeq (M'_\epsilon)_{s,s}$, then $\left(I - (M_\epsilon)_{s,s} \right)^{-1} \simeq \left(I - (M'_\epsilon)_{s,s} \right)^{-1}$.*

Proof. Since $(M_\epsilon)_{s,s} \simeq (M'_\epsilon)_{s,s}$, by Theorem 7.3,

$$
R \left((M_\epsilon)_{s,s} \right) = R \left((M'_\epsilon)_{s,s} \right) \quad \text{and} \quad C \left((M_\epsilon)_{s,s} \right) = C \left((M'_\epsilon)_{s,s} \right).
$$

Therefore, $R(M_\epsilon, \sigma) = R(M'_\epsilon, \sigma)$ for all $\sigma \in \mathcal{S}_n(s, s_i, s_j)$. This then implies that $\overline{\mathcal{P}}_{M_\epsilon}(s, s_i, s_j) = \overline{\mathcal{P}}_{M'_\epsilon}(s, s_i, s_j)$. Likewise, $C(M_\epsilon, \sigma) = C(M'_\epsilon, \sigma)$, $\forall \sigma \in \mathcal{S}_n(s, s_i, s_j)$. Therefore, by Theorem 7.25,

$$
R \left(I - (M_\epsilon)_{s,s} \right)^{-1} = R \left(I - (M'_\epsilon)_{s,s} \right)^{-1} \text{ and } C \left(I - (M_\epsilon)_{s,s} \right)^{-1} = C \left(I - (M'_\epsilon)_{s,s} \right)^{-1}
$$

so that, by Theorem 7.3, $\left(I - (M_\epsilon)_{s,s} \right)^{-1} \simeq \left(I - (M'_\epsilon)_{s,s} \right)^{-1}$. $\square$

We may now show that the reduction of a PMM is also a PMM and give formulas for its resistance and cost matrices in terms of minimum resistance paths.

Theorem 7.27. *Under the assumptions of Theorem 7.25, if, for each $\epsilon \succeq 0$, $\widehat{M}_\epsilon$ is the reduction of M_ϵ with respect to s, $\widehat{M}_\epsilon \in \mathrm{PMM}(|\overline{s}|)$ with*

$$R\left(\widehat{M}_\epsilon\right)_{i,j} = \min_{\sigma \in \mathcal{S}_n\left(s,\overline{s}_i,\overline{s}_j\right)} R\left(M_\epsilon, \sigma\right) \text{ and}$$

$$C\left(\widehat{M}_\epsilon\right)_{i,j} = \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}\left(s,\overline{s}_i,\overline{s}_j\right)} C\left(M_\epsilon, \sigma\right)$$

Proof. As before, s is open with respect to M_ϵ for sufficiently small ϵ, so the reduction is defined. By Theorem 5.6 c), $\widehat{M}_\epsilon = (M_\epsilon)_{\overline{s},\overline{s}} - (M_\epsilon)_{\overline{s},s} (\Lambda_\epsilon)_{s,s}^{-1} (M_\epsilon)_{s,\overline{s}}$. By Theorem 7.25, $-(\Lambda_\epsilon)_{s,s}^{-1} = \left(I - (M_\epsilon)_{s,s}\right)^{-1} \in \mathrm{Pert}$. Therefore, Theorem 7.6 implies that $\widehat{M}_\epsilon \in \mathrm{Pert}$. In addition, Corollary 5.11 guarantees that $\widehat{M}_\epsilon$ is unichain for $\epsilon \succeq 0$. Thus, $\widehat{M}_\epsilon \in \mathrm{PMM}(|\overline{s}|)$. The formulas for its resistance and cost matrices follow directly from Theorems 7.4 and 7.25. $\square$

Theorem 7.27 implies that reduction preserves asymptotic equivalence.

Corollary 7.28. *Given $M_\epsilon, M_\epsilon \in \mathrm{PMM}(n)$, $s \subset S_n$ open with respect to M_0 and $M_0{}'$ with $\left(\widehat{M}_\epsilon, p_\epsilon, \imath_\epsilon\right)$ and $\left(\widehat{M}'_\epsilon, p'_\epsilon, \imath'_\epsilon\right)$ the reductions with respect to s of M_ϵ and M'_ϵ, respectively, if $M_\epsilon \simeq M_\epsilon$, then $\widehat{M}_\epsilon \simeq \widehat{M}'_\epsilon$.*

Proof. Since $M_\epsilon \simeq M_\epsilon$, $R\left(M_\epsilon\right) = R\left(M'_\epsilon\right)$, which implies that

$$R\left(\widehat{M}_\epsilon\right)_{i,j} = \min_{\sigma \in \mathcal{S}_n\left(s,\overline{s}_i,\overline{s}_j\right)} R\left(M_\epsilon, \sigma\right) = \min_{\sigma \in \mathcal{S}_n\left(s,\overline{s}_i,\overline{s}_j\right)} R\left(M'_\epsilon, \sigma\right) = R\left(\widehat{M}'_\epsilon\right)_{i,j}$$

for any i, j. This implies that $\overline{\mathcal{P}}_{M_\epsilon}\left(s, \overline{s}_i, \overline{s}_j\right) = \overline{\mathcal{P}}_{M'_\epsilon}\left(s, \overline{s}_i, \overline{s}_j\right)$, for any i, j. Since $C\left(M_\epsilon\right) = C\left(M'_\epsilon\right)$, we also have

$$C\left(\widehat{M}_\epsilon\right)_{i,j} = \sum_{\sigma \in \overline{\mathcal{P}}_{M_\epsilon}\left(s,\overline{s}_i,\overline{s}_j\right)} C\left(M_\epsilon, \sigma\right) = \sum_{\sigma \in \overline{\mathcal{P}}_{M'_\epsilon}\left(s,\overline{s}_i,\overline{s}_j\right)} C\left(M'_\epsilon, \sigma\right) = C\left(\widehat{M}'_\epsilon\right)_{i,j}$$

as well. Thus, by Theorem 7.3, $\widehat{M}_\epsilon \simeq \widehat{M}'_\epsilon$. $\square$

Theorem 5.12 generalizes directly to PMMs.

Theorem 7.29. *Under the assumptions of Theorem 7.25, if* $\left(\widehat{M}_\epsilon, p_\epsilon, \imath_\epsilon\right)$ *is the reduction of* M_ϵ *with respect to* s, $\widehat{M}_\epsilon \sim_{\imath_\epsilon} M_\epsilon$.

Proof. Theorem 5.12 guarantees that $\widehat{M}_\epsilon \approx_{\imath_\epsilon} M_\epsilon$ for $\epsilon \succeq 0$. As in the proof of Theorem 7.15, $\imath_\epsilon \operatorname{stab}\left(\widehat{M}_\epsilon\right) = f(\epsilon)\operatorname{stab}(M_\epsilon)$ for some $f \in \mathcal{C}^+$, and we must show that $R(f) = 0$.

By Theorem 5.12, $\operatorname{stab}\left(\widehat{M}_\epsilon\right) = \pi_{\overline{s}}\imath_\epsilon \operatorname{stab}\left(\widehat{M}_\epsilon\right) = f(\epsilon)\pi_{\overline{s}}\operatorname{stab}(M_\epsilon)$. Therefore,

$$
\begin{aligned}
\operatorname{stab}\left(\widehat{M}_\epsilon\right)_i &= e_i^t \operatorname{stab}\left(\widehat{M}_\epsilon\right) = e_i^t f(\epsilon)\pi_{\overline{s}}\operatorname{stab}(M_\epsilon) \\
&= f(\epsilon)e_{\overline{s}_i}^t \operatorname{stab}(M_\epsilon) = f(\epsilon)\operatorname{stab}(M_\epsilon)_{\overline{s}_i}.
\end{aligned}
$$

Since $0 = R\left(\operatorname{stab}\left(\widehat{M}_\epsilon\right)\right)_i = R(f) + R\left(\operatorname{stab}(M_\epsilon)\right)_{\overline{s}_i}$, we must have $R(f) = 0$. $\square$

Theorem 7.29 implies that reduction preserves stochastic equivalence, as well.

Corollary 7.30. *Under the assumptions of Corollary 7.28, if* $M_\epsilon \sim M'_\epsilon$, *then* $\widehat{M}_\epsilon \sim \widehat{M}'_\epsilon$.

Proof. By Theorem 7.29, $\widehat{M}_\epsilon \sim_{\imath_\epsilon} M_\epsilon$ and $\widehat{M}'_\epsilon \sim_{\imath'_\epsilon} M'_\epsilon$. More specifically, $\imath_\epsilon \operatorname{stab}\left(\widehat{M}_\epsilon\right) = f(\epsilon)\operatorname{stab}(M_\epsilon)$ and $\imath_\epsilon \operatorname{stab}\left(\widehat{M}'_\epsilon\right) = f'(\epsilon)\operatorname{stab}(M'_\epsilon)$ for some $f, f' \in \mathcal{C}^+$ with $R(f) = R(f') = 0$. In fact, following the proof of Theorem 7.29, $\operatorname{stab}\left(\widehat{M}_\epsilon\right) = f(\epsilon)\pi_{\overline{s}}\operatorname{stab}(M_\epsilon)$ and $\operatorname{stab}\left(\widehat{M}'_\epsilon\right) = f'(\epsilon)\pi_{\overline{s}}\operatorname{stab}(M'_\epsilon)$. Moreover, since $J\operatorname{stab}\left(\widehat{M}_\epsilon\right) = 1 = J\operatorname{stab}\left(\widehat{M}'_\epsilon\right)$, $f(\epsilon) = \frac{1}{J\pi_{\overline{s}}\operatorname{stab}(M_\epsilon)}$ and $f'(\epsilon) = \frac{1}{J\pi_{\overline{s}}\operatorname{stab}(M'_\epsilon)}$. Since $M_\epsilon \sim M_\epsilon$, $\operatorname{stab}(M_\epsilon)_i \simeq \operatorname{stab}(M'_\epsilon)_i$ for all i, so that $\pi_{\overline{s}}\operatorname{stab}(M_\epsilon) \simeq \pi_{\overline{s}}\operatorname{stab}(M'_\epsilon)$. In particular, $f \simeq f'$, so that

$$
\operatorname{stab}\left(\widehat{M}_\epsilon\right) = f(\epsilon)\pi_{\overline{s}}\operatorname{stab}(M_\epsilon) \simeq f(\epsilon)'\pi_{\overline{s}}\operatorname{stab}(M'_\epsilon) = \operatorname{stab}\left(\widehat{M}'_\epsilon\right)
$$

$\square$

We now show that we may compute the reduction while only inverting a constant matrix.

Theorem 7.31. *Given* $M_\epsilon \in \operatorname{PMM}(n)$ *such that* $R(M_\epsilon)_{i,i} = 0$ *for all* i, *a communicating class,* $s' \subset S_n$, *of* M_0 *such that, for* $i \neq j \in s'$, $(M_\epsilon)_{i,j} = (M_0)_{i,j}$, *and a subset* $s \subset s'$ *such that* $\left|s' \setminus s\right| = 1$, *if* $\widehat{M}_\epsilon$ *is the reduction with respect to* s *of* M_ϵ, *there is a PMM,* $\widetilde{M}_\epsilon$ *such that* $\widehat{M}_\epsilon \sim \widetilde{M}_\epsilon$, *where*

$$
\left(\widetilde{M}_\epsilon\right)_{i,j} = \left((M_\epsilon)_{\overline{s},\overline{s}} + (M_\epsilon)_{\overline{s},s}\left(I - (M_0)_{s,s}\right)^{-1}(M_\epsilon)_{s,\overline{s}}\right)_{i,j}
$$

for $i \neq j$.

Proof. Notice that $(M_\epsilon)_{s,s} \simeq (M_0)_{s,s}$. Therefore, by Corollary 7.26,

$$\left(I - (M_\epsilon)_{s,s}\right)^{-1} \simeq \left(I - (M_0)_{s,s}\right)^{-1},$$

so that

$$\widehat{M_\epsilon} \simeq (M_\epsilon)_{\bar{s},\bar{s}} + (M_\epsilon)_{\bar{s},s} \left(I - (M_0)_{s,s}\right)^{-1} (M_\epsilon)_{s,\bar{s}}.$$

By Lemma 7.21, there is a PMM, $\widetilde{M_\epsilon}$, such that

$$\left(\widetilde{M_\epsilon}\right)_{i,j} = \left((M_\epsilon)_{\bar{s},\bar{s}} + (M_\epsilon)_{\bar{s},s} \left(I - (M_0)_{s,s}\right)^{-1} (M_\epsilon)_{s,\bar{s}}\right)_{i,j}$$

for $i \neq j$ so that $\widehat{M_\epsilon} \simeq \widetilde{M_\epsilon}$. By Theorem 7.10, $\widehat{M_\epsilon} \sim \widetilde{M_\epsilon}$. $\square$

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.

We begin by observing that, by Theorem 7.10, Theorem 7.11, and Corollary 7.30, 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.

Algorithm 2 To Compute the SSD of a PMM.

```
1: function v₀ = SSD (Mₑ) {
2:     C = commClasses (M₀);
3:     if (C.numClosed == 1)
4:          return(stab (M₀));
5:     if (C.nonTrivial > 0)
6:          (Mₑ, i, C) = reduce (Mₑ, C);
7:     (Mₑ, D) = nonUniformScale (Mₑ, C);
8:     return(normalize (iD (SSD (Mₑ))));
9: }
```

1: **function** $v_0 = \text{SSD}\,(M_\epsilon)$ {
2:      $C = \text{commClasses}\,(M_0)$;
3:      **if** $(C.\text{numClosed} == 1)$
4:         **return**$(\text{stab}\,(M_0))$;
5:      **if** $(C.\text{nonTrivial} > 0)$
6:         $(M_\epsilon, i, C) = \text{reduce}\,(M_\epsilon, C)$;
7:      $(M_\epsilon, D) = \text{nonUniformScale}\,(M_\epsilon, C)$;
8:      **return**$(\text{normalize}\,(iD\,(\text{SSD}\,(M_\epsilon))))$;
9: }

Algorithm 3 To Reduce a GSA.

1: **#define** uniformScale(M) (hasZeroOnDiagonalP(M) : $(I + M)/2$? M)
2: **function** $(M_\epsilon, i, C) = \text{reduce}\,(M_\epsilon, i, C)$ {
3:      $M_\epsilon = \text{uniformScale}\,(M_\epsilon)$;
4:      $M_\epsilon = \text{dropNonZeroR}\,(M_\epsilon, C)$;
5:      $i = \text{identityMatrix}\,(M_\epsilon.\text{dim}())$;
6:      **for** $(c = C.\text{first}(); c.\text{next}() \neq \text{NULL}; c = c.\text{next}())$ {
7:         $s = c.\text{members}().\text{rest}()$;
8:         $c.\text{setMembers}\,(c.\text{members}().\text{first}())$;
9:         $M_\epsilon = (M_\epsilon)_{\bar{s},\bar{s}} + (M_\epsilon)_{\bar{s},s} \left(I - (M_0)_{s,s}\right)^{-1} (M_\epsilon)_{s,\bar{s}}$;
10:         $i = iP_s \begin{pmatrix} I \\ \left(I - (M_0)_{s,s}\right)^{-1} (M_0)_{s,\bar{s}} \end{pmatrix}$;
11:      }
12:      **return**(M_ϵ, i, C);
13: }

Theorem 7.32. *Given* $M_\epsilon \in \mathrm{PMM}(n)$, *if* M_0 *is unichain, then* $\mathrm{stab}\, M_0 = \{\mathrm{ssd}\, M_\epsilon\}$.

Proof. Letting $v_\epsilon \equiv \mathrm{stab}\,(M_\epsilon)$, by definition, $M_\epsilon v_\epsilon = v_\epsilon$ for $\epsilon \succ 0$. Taking limits, we have $M_0 v_0 = v_0$, that is, $\mathrm{ssd}\, M_\epsilon = v_0 \in \mathrm{stab}\, M_0$. Since M_0 is unichain, $\mathrm{stab}\, M_0$ is a singleton, so that $\mathrm{stab}\, M_0 = \{\mathrm{ssd}\, M_\epsilon\}$. $\square$

Next, observe how **uniformScale** carries out the construction of Corollary 7.13 to guarantee that we have zero-resistance diagonal entries. Likewise, **dropNonZeroR** implements Lemma 7.22 by dropping all entries with non-zero resistance corresponding to edges within any communicating class in C. We may then appeal to Theorem 7.31 to justify 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, $\imath$. This is correct by Theorem 7.15. Also, 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.

Chapter 8

PMMs and Generalized Simulated Annealing

By restricting attention to a sequence $\epsilon_t = e^{-\frac{1}{T_t}}$ for some sequence $T_t \to 0$, for any given initial distribution, a perturbed Markov matrix defines an *inhomogenous* Markov chain, $\{X_t\}_{t=0}^{\infty}$, of a generalized simulated annealing (GSA) (Trouvé, 1996; Desai et al., 1994) with transition matrix $M^t = M_{\epsilon_t}$. Moreover, if $T_t \to 0$ slowly enough, Desai et al. (1994) describe the resulting process as "quasi-statically cooled". This is intended to connote that, in some sense, the limiting distribution of this process equals the limit of the stable distributions, v_{ϵ_t} as $t \to \infty$, i.e., the stochastically stable distribution of M_{ϵ}.

Specifically, assume that the resistance matrix, $R(M_{\epsilon})$, corresponds to the energy differences of a potential function, $U(i)$, so that $R(M_{\epsilon})_{i,j} = (U(i) - U(j))^+$. Notice that this implies that $R(M_{\epsilon})_{i,j} \neq \infty$ for all i and j. In particular, M_{ϵ} is *irreducible*, since all its entries are non-zero. For any such resistance matrix (with all finite entries), the corresponding potential function, U, is uniquely defined, up to an additive constant, and there is a unique choice with $\min_j U(j) = 0$.

Under certain conditions, the resistance matrix of the stable distribution of M_{ϵ} will allow us to recover this potential, i.e.,

$$R\left(\text{stab}\left(M_{\epsilon}\right)\right)_i = U(i) - \min_j U(j) \tag{8.1}$$

To be precise, Trouvé (1996) shows that Equation 8.1 holds iff Hajek's "weak reversibility" condition is satisfied. For example, it suffices for the communication matrix, $C(M_{\epsilon})$ to be

symmetric. Thus, the resistance, $R\left(\mathrm{stab}\left(M_\epsilon\right)\right)_i$, may also be called the *virtual energy*[1] at i in M_ϵ. By Theorem 7.4 a), we have that $\left(v_0\left(M_\epsilon\right)\right)_i > 0$ iff $R\left(\mathrm{stab}\left(M_\epsilon\right)\right)_i = 0$. Thus, the stochastically stable states of a PMM are also called its *ground* states.

In this chapter, we show how minor modifications to Algorithm 2 yields an efficient technique for computing $R\left(\mathrm{stab}\left(M_\epsilon\right)\right)$ when M_ϵ is irreducible. Based on the previous discussion, we will denote the set of all n-dimensional, irreducible PMMs as $\mathrm{GSA}(n)$. As before, we will show that:

- $R\left(\mathrm{stab}\left(M_\epsilon\right)\right)$ only depends on M_ϵ up to an equivalence relation (weaker than that of chapter 7) defined over its entries,

- the equivalence class of a GSA, M_ϵ, is determined by its resistance, $R\left(M_\epsilon\right)$,

- there are correspondingly weaker notions of equivalence and D-equivalence, which are preserved under the operations of scaling and reduction, and

- by appealing to the Markov Chain Tree Theorem, we again only need invert *constant* matrices in our constructions.

8.1 Equivalence of GSAs

As before, we begin by defining an equivalence relation on $\mathcal{C}^+$, generalizing it to Pert, and then specializing to $\mathrm{GSA}(n)$. First, define $f, g \in \mathcal{C}^+$ to be *asymptotic*, written as $f \simeq_w g$ iff $R(f) = R(g)$. The notation is motivated by fact that this is a strictly weaker notion than asymptotic equality. The terminology is justified by the following theorem.

Theorem 8.1. *If $f, g \in \mathcal{C}^+$, f is asymptotic to g iff $g(\epsilon) = 0 = f(\epsilon)$ for $\epsilon \succeq 0$ or $0 < \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} < \infty$.*

Proof. If f is asymptotic to g, there are two cases to consider. If $R(f) = R(g) = \infty$, then by Theorem 7.3 b), $f \simeq g \simeq 0$, so that $g(\epsilon) = 0 = f(\epsilon)$ for $\epsilon \succeq 0$. Otherwise, $r \equiv R(f) = R(g) < \infty$, so that $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} = 1 = \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{c'\epsilon^r}$ for $c, c' > 0$, and $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} = \frac{c}{c'} \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} \frac{c'\epsilon^r}{g(\epsilon)} = \frac{c}{c'}$, where $0 < \frac{c}{c'} < \infty$.

---

[1]Desai et al. (1994) call this the "stationary order" at i.

Conversely, if $g(\epsilon) = 0 = f(\epsilon)$ for $\epsilon \succeq 0$, then $R(f) = R(g) = \infty$. Otherwise, $0 < \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} < \infty$. This implies that $f \not\simeq 0$ and $g \not\simeq 0$. Since $f, g \in \mathcal{C}^+$, we then know that $\lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} = 1 = \lim_{\epsilon \to 0^+} \frac{g(\epsilon)}{c'\epsilon^{r'}}$ for $c, c' > 0$ and $0 \le r, r' < \infty$. Therefore, $1 = \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{c\epsilon^r} \lim_{\epsilon \to 0^+} \frac{c'\epsilon^{r'}}{g(\epsilon)} = \frac{c}{c'} \lim_{\epsilon \to 0^+} \frac{f(\epsilon)}{g(\epsilon)} \lim_{\epsilon \to 0^+} \epsilon^{r-r'}$, which implies that $R(f) = r = r' = R(g)$. $\square$

Likewise, for $M_\epsilon, M_\epsilon' \in \text{Pert}$, we will say that M_ϵ is *asymptotic* to M_ϵ', writing $M_\epsilon \simeq_w M'_\epsilon$, iff $(M_\epsilon)_{ij} \simeq_w (M'_\epsilon)_{ij}$ for all i, j. Theorem 7.6 implies that addition and multiplication of perturbed matrices is well-defined on equivalence classes under $\simeq_w$, as well. In the case of GSAs, we have the following definition.

Definition 8.2. *If* $M_\epsilon, M_\epsilon' \in \text{GSA}(n)$, $M_\epsilon \sim_w M_\epsilon'$ *iff* $\text{stab}(M_\epsilon) \simeq_w \text{stab}(M_\epsilon')$. *In this case, we say* M_ϵ *and* M_ϵ' *are* energy equivalent.

We now state the following easy consequences of Theorem 7.7.

Theorem 8.3.

a) *If* $f \simeq g$, *then* $f \simeq_w g$.

b) *If* $M_\epsilon \simeq M_\epsilon'$, *then* $M_\epsilon \simeq_w M_\epsilon'$.

c) *If* $M_\epsilon \sim M_\epsilon'$, *then* $M_\epsilon \sim_w M_\epsilon'$.

d) *If* $M_\epsilon \simeq_w M_\epsilon'$, *then* $M_\epsilon \sim_w M_\epsilon'$.

e) *If* $M_\epsilon \sim M_\epsilon'$ *and* $M_\epsilon \in \text{GSA}(n)$, *then* $M_\epsilon' \in \text{GSA}(n)$.

Proof. Part a) follows directly from Theorem 7.3. Part b) is an immediate consequence of part a), and implies part c). Part d) follows from Theorem 7.7 and Equation 7.1, as in the proof of Theorem 7.10.

To prove part e), notice that, by Corollary 5.15, since $\text{stab}(M_\epsilon) \simeq \text{stab}(M_\epsilon')$,

$$s = \{i \in S_n \mid (\text{stab}(M_\epsilon))_i \not\simeq 0\} = \{i \in S_n \mid (\text{stab}(M_\epsilon'))_i \not\simeq 0\} = s'$$

If $M_\epsilon \in \text{GSA}(n)$, then $s = S_n$, so that $s' = S_n$, M_ϵ is irreducible, and $M_\epsilon \in \text{GSA}(n)$. $\square$

We have the following analog of Theorem 7.12.

Theorem 8.4. *Given $f \in \mathcal{C}^+ \setminus 0$ and $M_\epsilon \in \mathrm{GSA}(n)$, such that*

a) $R(f) \leq R(M_\epsilon)_{i,j}$ *for all* $i \neq j$,

b) $(M_\epsilon)_{j,j} + f(\epsilon) - 1 \in \mathcal{C}^+$, *and*

c) $R(f) \leq R\left((M_\epsilon)_{j,j} + f(\epsilon) - 1\right)$ *for all* j,

if we define $M'_\epsilon = \frac{1}{f}(M_\epsilon - I) + I$, *then* $M'_\epsilon \in \mathrm{GSA}(n)$ *and* $M_\epsilon \sim_w M'_\epsilon$.

Proof. Since a GSA is a PMM, we may apply Theorem 7.12 to conclude that $M'_\epsilon \in \mathrm{PMM}(n)$ and $M_\epsilon \sim M'_\epsilon$. Theorem 8.3 e) then says that $M'_\epsilon \in \mathrm{GSA}(n)$ and Theorem 8.3 c) gives $M_\epsilon \sim_w M'_\epsilon$. $\square$

As before, Corollary 7.13 and Theorem 8.3 allow us to restrict attention to GSAs with $R(M_\epsilon)_{j,j} = 0$, for all j.

Corollary 8.5. *Given* $M_\epsilon \in \mathrm{GSA}(n)$, *if we define* $M'_\epsilon = \frac{1}{2}(M_\epsilon - I) + I$, *then* $M'_\epsilon \in \mathrm{GSA}(n)$, $M_\epsilon \sim_w M'_\epsilon$, *and* $R\left(M'_\epsilon\right)_{j,j} = 0$, *for all* $j \in S_n$.

Minor modifications to Lemma 7.16 give the corresponding result for GSAs.

Lemma 8.6. *Given* $M_\epsilon \in \mathrm{GSA}(n)$ *with* $n > 1$, *such that all communicating classes of* M_0 *are singletons, define* $s \equiv \min_{j \notin \mathcal{T}} \min_{i \neq j} R(M_\epsilon)_{i,j}$, *where* $\mathcal{T}$ *is the set of transient states of* M_0. *Likewise, let* $c \equiv 2 \max_{j \notin \mathcal{T}} \sum_{s=R(M_\epsilon)_{i,j}} C(M_\epsilon)_{i,j}$. *Then* $0 < s < \infty$ *and* $c > 0$ *so we may define* $f(\epsilon) \equiv c\epsilon^s$ *and*

$$(i_\epsilon)_{i,j} \equiv \begin{cases} 0 & \text{if } i \neq j \\ f(\epsilon) & \text{if } i = j \text{ and } j \in \mathcal{T} \\ 1 & \text{otherwise} \end{cases}$$

so that $i_\epsilon \in \mathrm{Pert}(n)$ *and* $M'_\epsilon = \frac{1}{f}\left(\overline{M}_\epsilon - I\right) + I$ *is energy* i_ϵ-*equivalent to* M_ϵ. *In particular,*

$$R\left(\mathrm{stab}\left(M_\epsilon\right)\right)_j = \begin{cases} s + R\left(M'_\epsilon\right)_j & \text{if } j \in \mathcal{T} \\ R\left(M'_\epsilon\right)_j & \text{otherwise} \end{cases}.$$

Proof. Since any GSA is a PMM, the proof of Lemma 7.16 applies unchanged, except for showing that $s < \infty$. Since M_ϵ is irreducible, there are paths in $G_-(M_\epsilon)$ from any vertex to every other vertex. In particular, each closed indexa has at least one outgoing edge, so that $s < \infty$. Continuing the proof of Lemma 7.16, we see that M_ϵ' is i_ϵ-equivalent to M_ϵ. In other words, $g(\epsilon)\operatorname{stab}(M_\epsilon) = i_\epsilon \operatorname{stab}(M_\epsilon')$ for some $g \in \mathcal{C}^+$ with $R(g) = 0$. Therefore, by Theorem 7.6, $R(\operatorname{stab}(M_\epsilon)) = R(g) + R(\operatorname{stab}(M_\epsilon)) = R(i_\epsilon \operatorname{stab}(M_\epsilon'))$. Applying Theorem 7.6 again, the $R\left(i_\epsilon \operatorname{stab}(M_\epsilon')\right)_j = \min_{k \in S_m} \left\{ R(i_\epsilon)_{j,k} + R(\operatorname{stab}(M_\epsilon'))_k \right\}$. Since i_ϵ is diagonal, $R(i_\epsilon)_{j,k} = \infty$, unless $j = k$, so that the minimum occurs when $j = k$ and $R\left(i_\epsilon \operatorname{stab}(M_\epsilon')\right)_j = R(i_\epsilon)_{j,j} + R(\operatorname{stab}(M_\epsilon'))_k$. Since $R(i_\epsilon)_{j,j} = s$, if $j \in \mathcal{T}$, and 0, otherwise, the result follows. $\square$

By repeatedly applying Lemma 8.6, 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 8.7. *Given $M_\epsilon \in \mathrm{GSA}(n)$ with $n > 1$, there exists a diagonal $i_\epsilon \in \mathrm{Pert}(n)$ and $M_\epsilon' \in \mathrm{GSA}(n)$ such that $M_\epsilon \sim_{i_\epsilon} M_\epsilon'$ and M_0' possesses a non-trivial communicating class.*

Proof. As before, since any GSA is a PMM, we may repeat the proof of Corollary 7.17. In other words, we proceed by repeatedly applying Lemma 8.6, so that, at each step, $M_\epsilon' = (M_\epsilon)_{D_\epsilon}$, for a diagonal matrix, D_ϵ. We know that $M_\epsilon' \in \mathrm{PMM}(n)$. However, scaling by a diagonal matrix only affects the magnitude of the edge weights. In particular, it can only introduce or eliminate self-loops, so that $\mathcal{P}(M_\epsilon') = \mathcal{P}(M_\epsilon)$. Thus, since M_ϵ is irreducible, $\mathcal{P}(M_\epsilon)$ and $\mathcal{P}(M_\epsilon')$ are complete, so that M_ϵ' is irreducible, i.e., $M_\epsilon' \in \mathrm{GSA}(n)$. As before, $G_-(M_0') \supset G_-(M_0)$ is a strictly larger graph, so this is guaranteed to terminate by the time $G_-(M_0')$ is complete, if not sooner. $\square$

8.2 Reduction of GSAs

Notice that, since a GSA is a PMM, Theorem 7.27 gives us the corresponding result for GSAs almost immediately.

Theorem 8.8. *For any $M_\epsilon \in GSA(n)$ and $s \subset S_n$ an open set with respect to M_0, if, for each $\epsilon \succeq 0$, $\widehat{M_\epsilon}$ is the reduction of M_ϵ with respect to s, $\widehat{M_\epsilon} \in \mathrm{GSA}(|\overline{s}|)$ with*

$$R\left(\widehat{M_\epsilon}\right)_{i,j} = \min_{\sigma \in \mathcal{S}_n\left(s, \overline{s}_i, \overline{s}_j\right)} R(M_\epsilon, \sigma)$$

Proof. By Theorem 7.27, $\widehat{M}_\epsilon \in \mathrm{PMM}(|\bar{s}|)$ and Corollary 5.11 then implies that $\widehat{M}_\epsilon$ is irreducible for $\epsilon \succeq 0$, i.e., $\widehat{M}_\epsilon \in \mathrm{GSA}(|\bar{s}|)$. $\square$

This implies a result corresponding to Corollary 7.28 for PMMs.

Corollary 8.9. *Given* $M_\epsilon, M_\epsilon \in \mathrm{GSA}(n)$, $s \subset S_n$ *open with respect to* M_0 *and* $M_0{}'$ *with* $\left(\widehat{M}_\epsilon, p_\epsilon, \imath_\epsilon\right)$ *and* $\left(\widehat{M'_\epsilon}, p'_\epsilon, \imath'_\epsilon\right)$ *the reductions with respect to* s *of* M_ϵ *and* M'_ϵ, *respectively, if* $M_\epsilon \simeq_w M_\epsilon$, *then* $\widehat{M}_\epsilon \simeq_w \widehat{M'_\epsilon}$.

Notice that the first part of Theorem 7.18 may be restated for GSAs, as follows.

Theorem 8.10. *Given* $M_\epsilon, M'_\epsilon \in \mathrm{GSA}(n)$, *if*

a) $\overline{T}\left(M_\epsilon, i\right) = \overline{T}\left(M'_\epsilon, i\right)$ *and*

b) $R\left(M_\epsilon\right)_{\sigma(j),j} = R\left(M'_\epsilon\right)_{\sigma(j),j}, \forall j \neq i, \sigma \in \overline{T}\left(M_\epsilon, i\right)$

for every $1 \leq i \leq n$, *then* $M_\epsilon \sim_w M'_\epsilon$.

We may use Theorem 8.10 to obtain an analog of Theorem 7.31 for GSAs, guaranteeing that we need only invert a constant matrix when computing the reduction of a GSA. To prove this, we must first prove the following analog of Lemma 7.22.

Lemma 8.11. *Given* $M_\epsilon \in \mathrm{GSA}(n)$, *if* $0 < R\left(M_\epsilon\right)_{i,j}, R\left(M_\epsilon\right)_{j,k} < \infty$, $R\left(M_\epsilon\right)_{i,k} \geq R\left(M_\epsilon\right)_{i,j} + R\left(M_\epsilon\right)_{j,k}$, *and* $M'_\epsilon \equiv M_\epsilon - (M_\epsilon)_{i,k}\left(e_i e_k^t - e_k e_k^t\right)$, *then* $M'_\epsilon \in \mathrm{GSA}(n)$ *and* $M_\epsilon \sim_w M'_\epsilon$.

Proof. Observe that M'_ϵ is just M_ϵ with the $(i,k)^{\text{th}}$ entry set to 0, adding it to the diagonal. Thus, $M'_\epsilon \in \mathrm{Pert}$. Likewise, $G\left(M'_\epsilon\right)$ is the same as $G\left(M_\epsilon\right)$, except for the removal of the edge from v_k to v_i. However, since $G\left(M_\epsilon\right)$ contains a path of length 2 from v_k to v_i, $\mathcal{P}\left(M_\epsilon\right) = \mathcal{P}\left(M'_\epsilon\right)$. In particular, M'_ϵ is irreducible, and $M'_\epsilon \in \mathrm{GSA}(n)$.

We will show that if $\sigma \in T\left(M_\epsilon, r\right)$ is a spanning subtree rooted at r with $\sigma(k) = i$, then there is another spanning subtree, $\sigma \in T\left(M_\epsilon, r\right)$, rooted at r with $\sigma'(k) = j$, and $R\left(M_\epsilon, \sigma'\right) \leq R\left(M_\epsilon, \sigma\right)$. Thus, when considering minimum resistance spanning subtrees, we may restrict attention to those that do not contain an edge from v_k to v_i. Since the resistance of such subtrees are necessarily equal for M_ϵ and M'_ϵ, Theorem 8.10 would then imply that $M_\epsilon \sim_w M'_\epsilon$.

Assume we are given $\sigma \in T\left(M_\epsilon, s\right)$ 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 v_i and v_r are in $\mathcal{T}_1$ and v_k is in $\mathcal{T}_2$. Now v_j must be in one of these subtrees. If v_j is in $\mathcal{T}_1$, by adding the edge from v_k to v_j we obtain a tree with total resistance which has decreased by $R\left(M_\epsilon\right)_{i,j}$. Formally, defining $\sigma'(t) = \sigma(t)$, for all $t \neq k$, and $\sigma'(k) = j$, $R\left(M_\epsilon, \sigma'\right) = R\left(M_\epsilon, \sigma\right) - R\left(M_\epsilon\right)_{i,j} < R\left(M_\epsilon, \sigma\right)$.

Otherwise, v_j is in $\mathcal{T}_2$. Deleting the edge from v_j in $\mathcal{T}_2$ splits it into two smaller trees, $\mathcal{T}_2'$ and $\mathcal{T}_2''$, where we may assume that v_j is in $\mathcal{T}_2''$ and v_k is in $\mathcal{T}_2'$. By adding the edges from v_k to v_j and from v_j to v_i, we obtain a tree with total resistance no greater than before, but which does not include the edge from v_k to v_i. Formally, defining $\sigma'(t) = \sigma(t)$, for all $t \neq k, j$, $\sigma'(k) = j$, $\sigma'(j) = i$, $R\left(M_\epsilon, \sigma'\right) \leq R\left(M_\epsilon, \sigma\right) - R\left(M_\epsilon\right)_{i,k} + R\left(M_\epsilon\right)_{i,j} + R\left(M_\epsilon\right)_{j,k} \leq R\left(M_\epsilon, \sigma\right)$. $\square$

Theorem 8.12. *Given $M_\epsilon \in \mathrm{GSA}(n)$ such that $R\left(M_\epsilon\right)_{i,i} = 0$ for all i, a communicating class, $s' \subset S_n$, of M_0 such that, for $i \neq j \in s'$, $\left(M_\epsilon\right)_{i,j} = \left(M_0\right)_{i,j}$, and a subset $s \subset s'$ such that $\left|s' \setminus s\right| = 1$, if $\left(\widehat{M_\epsilon}, p_\epsilon, \imath_\epsilon\right)$ is the reduction with respect to s of M_ϵ, there is an $M_\epsilon' \in \mathrm{GSA}(|\bar{s}|)$ such that $\left(M_\epsilon'\right)_{i,j} = \left(p_0\left(M_\epsilon - I\right)\imath_0 + I\right)_{i,j}$ for $i \neq j$ and $\widehat{M_\epsilon} \sim_w M_\epsilon'$.*

Proof. By Theorem 5.8 c),

$$\left(\widehat{M_\epsilon}\right)_{i,j} = \sum_{\sigma \in \mathcal{P}_{M_\epsilon}\left(s, \bar{s}_i, \bar{s}_j\right)} W\left(M_\epsilon, \sigma\right)$$

For any i and j for which $0 < R\left(\widehat{M_\epsilon}\right)_{i,j} < \infty$, choose $\sigma \in \mathcal{P}_{M_\epsilon}\left(s, \bar{s}_i, \bar{s}_j\right)$ such that $|\sigma| \equiv l$ and $R\left(M_\epsilon, \sigma\right)$ are minimum.

If $l = 1$, then we replace $\left(\widehat{M_\epsilon}\right)_{i,j}$ by $\left(M_\epsilon\right)_{\bar{s}_i, \bar{s}_j}$, adding the difference to $\left(\widehat{M_\epsilon}\right)_{j,j}$ to obtain an asymptotically equivalent GSA. Likewise, if $l > 1$ and both $R\left(M_\epsilon\right)_{\sigma_l, \sigma_{l-1}} > 0$ and $R\left(M_\epsilon\right)_{\sigma_1, \sigma_0} > 0$, we will show that, by Lemma 8.11, we may set $\left(\widehat{M_\epsilon}\right)_{i,j}$ to 0, adjusting $\left(\widehat{M_\epsilon}\right)_{j,j}$, to obtain to obtain an asymptotically equivalent GSA. Repeating these two operations as many times as necessary, we obtain $\widetilde{M_\epsilon} \in \mathrm{GSA}\left(\bar{s}\right)$ such that $\widetilde{M_\epsilon} \sim_w \widehat{M}$, and $0 = R\left(\widetilde{M_\epsilon}\right)_{i,j}$ or $\infty = R\left(\widetilde{M_\epsilon}\right)_{i,j}$ or $R\left(\widetilde{M_\epsilon}\right)_{i,j}$ is that of any minimum resistance walk with exactly one non-zero resistance edge.

To prove that the conditions of Lemma 8.11 hold, let $s' \setminus s = \{q\}$ and assume that $r_1 \equiv R\left(M_\epsilon\right)_{\sigma_l, \sigma_{l-1}} > 0$ and $r_2 \equiv R\left(M_\epsilon\right)_{\sigma_1, \sigma_0} > 0$. Since q is in the same communicating

class as σ_{l-1} and σ_1, there exist walks in s from q to σ_{l-1} and σ_1, respectively. Adding the two non-zero resistance edges gives walks from q to $\sigma_l = i$ and from $\sigma_0 = j$ to q through s, with resistance r_1 and r_2, respectively. Thus, $r_1 \geq R\left(\widetilde{M}_\epsilon\right)_{i,q}$ and $r_2 \geq R\left(\widetilde{M}_\epsilon\right)_{q,j}$, so that $R\left(M_\epsilon, \sigma\right)_{i,j} = r_1 + r_2 \geq R\left(\widetilde{M}_\epsilon\right)_{i,q} + R\left(\widetilde{M}_\epsilon\right)_{q,j}$, and the conditions of Lemma 8.11 are satisfied.

Now notice that, if we let $\overline{M}_\epsilon \equiv p_0\left(M_\epsilon - I\right)\imath_0 + I$ that $\overline{M}_0 = \widehat{M}_0$. In particular, $\overline{M}_0$ is Markov. Since $R\left(M_\epsilon\right)_{i,i} = 0$ for all i, by Theorem 8.8, $R\left(\widehat{M}_\epsilon\right)_{i,i} = 0$, so that $\widehat{M}_0 > 0$, and hence, for all i, $\left(\overline{M}_\epsilon\right)_{i,i} > 0$ for $\epsilon \succeq 0$. Therefore, $\sum_{i \neq j}\left(\overline{M}_0\right)_{i,j} = 1 - \left(\overline{M}_\epsilon\right)_{i,i} < 1$.

We wish to show that $\left(\overline{M}_\epsilon\right)_{i,j} \in \mathcal{C}^+$ for $i \neq j$. By definition,

$$
\begin{aligned}
\overline{M}_\epsilon &= p_0\left(M_\epsilon - I\right)\imath_0 + I \\
&= \left(\begin{array}{cc} I & -(M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1} \end{array}\right) P_s^t \Lambda_\epsilon P_s \left(\begin{array}{c} I \\ -(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \end{array}\right) + I \\
&= \left(\begin{array}{cc} I & -(M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1} \end{array}\right) \left(\begin{array}{cc} (\Lambda_\epsilon)_{\overline{s},\overline{s}} & (M_\epsilon)_{\overline{s},s} \\ (M_\epsilon)_{s,\overline{s}} & (\Lambda_\epsilon)_{s,s} \end{array}\right) \left(\begin{array}{c} I \\ -(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \end{array}\right) + I \\
&= \left(\begin{array}{cc} I & -(M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1} \end{array}\right) \left(\begin{array}{c} (\Lambda_\epsilon)_{\overline{s},\overline{s}} - (M_\epsilon)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \\ (M_\epsilon)_{s,\overline{s}} - (\Lambda_\epsilon)_{s,s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \end{array}\right) + I \\
&= (\Lambda_\epsilon)_{\overline{s},\overline{s}} - (M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_\epsilon)_{s,\overline{s}} - (M_\epsilon)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \\
&\quad + (M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(\Lambda_\epsilon)_{s,s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} + I \\
&= (M_\epsilon)_{\overline{s},\overline{s}} - (M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_\epsilon)_{s,\overline{s}} - (M_\epsilon)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} \\
&\quad + (M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(M_\epsilon)_{s,s}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}} - (M_0)_{\overline{s},s}\,(\Lambda_0)_{s,s}^{-1}\,(\Lambda_0)_{s,s}^{-1}\,(M_0)_{s,\overline{s}}
\end{aligned}
$$

Focusing attention on the off-diagonal entries, by Corollary 4.5, we may rewrite this in

terms of walks, as follows. For $i \neq j$,

$$
\begin{aligned}
\left(\overline{M}_\epsilon\right)_{i,j} &= \left(M_\epsilon\right)_{\overline{s}_i,\overline{s}_j} + \sum_k \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} W\left(M_0,\sigma\right)\left(M_\epsilon\right)_{s_k,\overline{s}_j} \\
&\quad + \sum_k \sum_{\sigma \in \mathcal{P}_{M_0}\left(s,s_k,\overline{s}_j\right)} \left(M_\epsilon\right)_{\overline{s}_i,s_k} W\left(M_0,\sigma\right) \\
&\quad + \sum_{k,k'} \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} \sum_{\sigma' \in \mathcal{P}_{M_0}\left(s,s_{k'},\overline{s}_j\right)} W\left(M_0,\sigma\right)\left(M_\epsilon\right)_{s_k,s_{k'}} W\left(M_0,\sigma'\right) \\
&\quad - \sum_k \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} \sum_{\sigma' \in \mathcal{P}_{M_0}\left(s,s_k,\overline{s}_j\right)} W\left(M_0,\sigma\right) W\left(M_0,\sigma'\right) \qquad (8.2)
\end{aligned}
$$

If $R\left(\widehat{M}_\epsilon\right)_{i,j} = \infty$, then there are no walks in $G\left(M_\epsilon\right)$ from $\overline{s}_j$ to $\overline{s}_i$ through s, so that all the terms in Equation 8.2 vanish as well, i.e., $\left(\overline{M}_\epsilon\right)_{i,j} = 0$ for $\epsilon \succeq 0$. If $R\left(\widehat{M}_\epsilon\right)_{i,j} = 0$, since $\left(\overline{M}_0\right)_{i,j} = \left(\widehat{M}_0\right)_{i,j} > 0$, $\left(\overline{M}_\epsilon\right)_{i,j} > 0$ for $\epsilon \succeq 0$. Finally, if $0 < R\left(\widehat{M}_\epsilon\right)_{i,j} < \infty$, then there exists at least one walk in $G\left(M_\epsilon\right)$ from $\overline{s}_j$ to $\overline{s}_i$ through s, and every walk contains at least one edge with non-zero resistance. In this case,

$$
\begin{aligned}
\left(\overline{M}_\epsilon\right)_{i,j} &= \left(M_\epsilon\right)_{\overline{s}_i,\overline{s}_j} + \sum_k \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} W\left(M_0,\sigma\right)\left(M_\epsilon\right)_{s_k,\overline{s}_j} \\
&\quad + \sum_k \sum_{\sigma \in \mathcal{P}_{M_0}\left(s,s_k,\overline{s}_j\right)} \left(M_\epsilon\right)_{\overline{s}_i,s_k} W\left(M_0,\sigma\right) \qquad (8.3) \\
&\quad + \sum_{k,k'} \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} \sum_{\sigma' \in \mathcal{P}_{M_0}\left(s,s_{k'},\overline{s}_j\right)} W\left(M_0,\sigma\right)\left(M_\epsilon\right)_{s_k,s_{k'}} W\left(M_0,\sigma'\right) \geq 0
\end{aligned}
$$

since

$$
\sum_k \sum_{\sigma \in \mathcal{P}_{M_0}(s,\overline{s}_i,s_k)} \sum_{\sigma' \in \mathcal{P}_{M_0}\left(s,s_k,\overline{s}_j\right)} W\left(M_0,\sigma\right) W\left(M_0,\sigma'\right)
$$

is a sum over zero-resistance walks. Moreover, $R\left(\overline{M}_\epsilon\right)_{i,j}$ is the minimum resistance taken over all walks from $\overline{s}_j$ to $\overline{s}_i$ through s with exactly one non-zero resistance edge.

So far we have shown that $\left(\overline{M}_\epsilon\right)_{i,j} \in \mathcal{C}^+$ for all $i \neq j$ and that $R\left(\overline{M}_\epsilon\right)_{i,j} = R\left(\widetilde{M}_\epsilon\right)_{i,j}$ for all $i \neq j$. Consequently, $G_-\left(\overline{M}_\epsilon\right)$ and $G_-\left(\widetilde{M}_\epsilon\right)$ only differ in self-loops, so that $\mathcal{P}\left(\overline{M}_\epsilon\right) = \mathcal{P}\left(\widetilde{M}_\epsilon\right)$. Therefore, $G_-\left(\overline{M}_\epsilon\right)$ is unichain and Lemma 7.21 gives an $M'_\epsilon \in$ PMM(n), such that $\left(M'_\epsilon\right)_{i,j} = \left(\overline{M}_\epsilon\right)_{i,j}$ for $i \neq j$. In particular, M'_ϵ is irreducible, and hence $M'_\epsilon \in$ GSA(n).

Moreover, $R\left(M'_\epsilon\right)_{i,j} = R\left(\overline{M}_\epsilon\right)_{i,j} = R\left(\widetilde{M}_\epsilon\right)_{i,j}$ for $i \neq j$, and $R\left(M'_\epsilon\right)_{i,i} = 0 = R\left(\widetilde{M}_\epsilon\right)_{i,i}$ for all i. Therefore, $M'_\epsilon \approx_w \widetilde{M}_\epsilon$ and $M'_\epsilon \sim_w \widetilde{M}_\epsilon \sim_w \widehat{M}_\epsilon$. $\square$

We these results, Algorithm 2 carries almost unchanged for GSAs. However, using naturality of the reduce construction (i.e., Theorem 5.33), unlike the case of PMMs, we may reduce all communicating classes simultaneously.

Theorem 8.13. *Given $M_\epsilon \in \mathrm{GSA}(n)$ such that $R\left(M_\epsilon\right)_{i,i} = 0$ for all i, and for $i \neq j$ in the same communicating class of M_0, $(M_\epsilon)_{i,j} = (M_0)_{i,j}$, and a subset $s \subset S_n$ containing all but one representative of each communicating class in M_0, if $\left(\widehat{M}_\epsilon, p_\epsilon, \imath_\epsilon\right)$ is the reduction with respect to s of M_ϵ, there is an $M'_\epsilon \in \mathrm{GSA}(|\overline{s}|)$ such that $\left(M'_\epsilon\right)_{i,j} = \left(p_0\left(M_\epsilon - I\right)\imath_0 + I\right)_{i,j}$ for $i \neq j$ and $\widehat{M}_\epsilon \sim_w M'_\epsilon$.*

Proof. Assume that M_0 has k non-trivial (i.e., containing more than one element) communicating classes, and $\left(\widehat{M}_{j,\epsilon}, p_{j,\epsilon}, \imath_{j,\epsilon}\right)$ is the result of eliminating all but the chosen member of the jth communicating class from $\widehat{M}_{j-1,\epsilon}$, where $\widehat{M}_{0,\epsilon} \equiv M_\epsilon$. By Theorem 5.36, $\widehat{M}_{k,\epsilon} = \widehat{M}_\epsilon$ and $\widehat{M}_{j,\epsilon} = p_{j,\epsilon}\ldots p_{1,\epsilon}\left(M_\epsilon - I\right)i_{1,\epsilon}\ldots i_{j,\epsilon} + I$. Applying Theorem 8.12 k times gives that

$$
\begin{aligned}
\widehat{M}_\epsilon \quad &\sim_w \quad p_{k,0}p_{k-1,\epsilon}\ldots p_{1,\epsilon}\left(M_\epsilon - I\right)i_{1,\epsilon}\ldots i_{k-1,\epsilon}i_{k,0} + I \sim_w \cdots \\
&\sim_w \quad p_{k,0}\ldots p_{1,0}\left(M_\epsilon - I\right)i_{1,0}\ldots i_{k,0} + I
\end{aligned}
$$

which, by Theorem 5.36 again, is $\widehat{\widehat{M}}_\epsilon = p_0\left(M_\epsilon - I\right)i_0 + I$. $\square$

8.3   The Energy Algorithm

In this section, we present our algorithm for computing the potential energy of a GSA and prove that it is correct. The intuition and proof of correctness is generally the same as for Algorithm 2.

As before, we may replace M_ϵ by an energy equivalent GSA without affecting the final result. In particular, we may represent M_ϵ just by resistance matrix, using Theorem 7.6 to carry out subsequent algebraic operations. Notice for example that **uniformScale** simply replaces each diagonal entry of the resistance matrix with a 0. As before, Lemma 8.6 and Corollary 8.7 guarantee that we either terminate immediately or we reduce the dimension of M_ϵ, guaranteeing that we will terminate eventually.

Algorithm 4 To Compute the Energy of a GSA.

1: **function** $r =$ Energy (M_ϵ) {
2:     $C =$ commClasses (M_0);
3:     **if** $(\dim (M_\epsilon) == 1)$
4:         **return**$((0))$;
5:     **if** $(C.\text{nonTrivial} > 0)$
6:         $(M_\epsilon, i, C) =$ reduce (M_ϵ, C);
7:     $(M_\epsilon, D) =$ nonUniformScale (M_ϵ, C);
8:     **return**$(iD (\text{Energy} (M_\epsilon)))$;
9: }

Algorithm 5 To Reduce a GSA.

1: #**define**   uniformScale(M) $(\text{hasZeroOnDiagonalP} (M) \ : \ (I + M)/2 \ ? \ M)$
2: **function** $(M_\epsilon, i, C) =$ reduce (M_ϵ, i, C) {
3:     $M_\epsilon =$ uniformScale (M_ϵ);
4:     $M_\epsilon =$ dropNonZeroR (M_ϵ, C);
5:     $(s, C) =$ chooseMaximal (C);
6:     $p = \left(\begin{array}{cc} I & (M_0)_{\overline{s},s} \left(I - (M_0)_{s,s} \right)^{-1} \end{array} \right) P_s^t$;
7:     $i = P_s \left(\begin{array}{c} I \\ \left(I - (M_0)_{s,s} \right)^{-1} (M_0)_{s,\overline{s}} \end{array} \right)$;
8:     $M_\epsilon = p (M_\epsilon - I)_{s,s} i + I$;
9:     **return**(M_ϵ, i, C);
10: }

Besides the different termination condition, the other key difference between Algorithms 4 and 2 is the way in which we compute the reduction, while only inverting constant matrices. We begin Algorithm 5, as before, by guaranteeing that we have zero-resistance entries on the diagonal and within any submatrix corresponding to a communicating class. We then use **chooseMaximal** tp choose a set of representatives from each communicating class, returning their complement as s and the corresponding singleton sets as C. We may then use Theorem 8.12 to guarantee that applying the corresponding quotient operators of M_0 to M_ϵ yields a GSA which is energy equivalent to the quotient of M_ϵ with respect to s. This is essentially the algorithm of Gambin and Pokarowski (2001). However, their algorithm is mainly combinatorial, operating primarily as a recursive algorithm on graphs. Since our algorithm recursively operates on GSAs, we believe that it is conceptually more satisfying.

Part III

Two Related Algorithms

Chapter 9

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, $G = (V, E, d, s, t)$, where the vertices, V, correspond to the individuals being ranked, the edges, E, correspond to the set of comparisons, and $s(\alpha), t(\alpha) : E \to V$ 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, $x(t(\alpha)) - x(s(\alpha)) = d(\alpha), \ \forall \, \alpha \in E$, 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 9.3.

For a ranking to be consistent with the data, we must at least have $0 \leq x\left(t(\alpha)\right) - x\left(s(\alpha)\right) \leq d(\alpha), \ \forall\, \alpha \in E$. 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 $x\left(t(\alpha)\right) - x\left(s(\alpha)\right) \leq d(\alpha), \ \forall\, \alpha \in E$. 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 $(s,t) = (s(\alpha), t(\alpha))$ 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:

$$C(G) = \left\{ x : V \to \mathbb{R} \ \middle|\ 0 \leq x\left(t(\alpha)\right) - x\left(s(\alpha)\right) \leq d(\alpha), \ \forall\, \alpha \in E \right\}.$$

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 $x\left(t(\alpha)\right) - x\left(s(\alpha)\right) \leq x'\left(t(\alpha)\right) - x'\left(s(\alpha)\right), \forall\, \alpha \in E$. 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 $\sum_{\alpha \in E} x(t(\alpha)) - x(s(\alpha))$.

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 $x(t(\alpha)) - x(s(\alpha)) = d(\alpha)$ correspond to a choice of edges, while the remaining equality constraints, $x(t(\alpha)) = x(s(\alpha))$ 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 9.2 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.

9.1   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 $s \rightsquigarrow v_0 \leftsquigarrow v_1 \rightsquigarrow \cdots \leftsquigarrow v_k \rightsquigarrow t$. 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 $d(\alpha)$ 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 9.1. *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, α, we have $x(t(\alpha)) - x(s(\alpha)) \leq d(\alpha)$. For each reversed edge, we have $0 \leq x(t(\alpha)) - x(s(\alpha))$, so that $x(s(\alpha)) - x(t(\alpha)) \leq 0$. 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 $x(s) \leq x(t)$. $\square$

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,

$$
\begin{aligned}
I(G) &= \{(s, t) \mid \deg_{in}(s) = \deg_{out}(t) = 0\} \quad \text{or} \\
I(G) &= \{(s, t) \mid s = s(\alpha), t = t(\alpha), \alpha \in E\}.
\end{aligned}
$$

Lemma 9.2. *If $x' - x$ is constant, then $x \sim x'$. Moreover, if $x \in C(G)$, then $x' \in C(G)$.*

Proof. If $x' = x + c$, $x'(t') - x'(s') = x(t') - x(s'), \forall s', t' \in V$. Since $\sim$ and $C(G)$ are defined solely in terms collections of such differences, the conclusions of the Lemma follow. $\square$

Theorem 9.3. $\max C(G) \neq \emptyset$.

Proof. Let $M = \max_{s,t \in V} l_G(s, t)$. By Lemma 9.1 $-l_G(v, u) \leq x(u) - x(v) \leq l_G(u, v)$,

so that $|x(u) - x(v)| \leq \max \{l_G(u,v), l_G(v,u)\} \leq M$. By Lemma 9.2, 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 $D(x) < D(x')$, which is a contradiction. $\square$

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 9.4. *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 9.1. Likewise, $x(t) \leq x(s)$, so that $x(s) = x(t)$ for s and t within the same strongly connected component. $\square$

9.2   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, $G = (V, E, d, s, t)$, let $E_0 = \{\alpha \in E \mid d(\alpha) = 0\}$ be the set of 0 weight edges. We may define a new graph $T(G) = (V', E', d', s', t')$ 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 $d'(\alpha, 0) \equiv d(\alpha), d'(\alpha, 1) \equiv 0$, $s'(\alpha, 0) \equiv s(\alpha), s'(\alpha, 1) \equiv t(\alpha), e'(\alpha, 0) \equiv t(\alpha)$, and $e'(\alpha, 1) \equiv s(\alpha)$. Notice that $C(G) = C(T(G))$, since we are simply adding redundant equality constraints.

Theorem 9.5. *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 $0 \leq x(t(\alpha)) - x(s(\alpha)) \leq d(\alpha)$. Since $s'(\alpha, 0) = s(\alpha), t'(\alpha, 0) = t(\alpha)$, and $d'(\alpha, 0) = d(\alpha), 0 \leq x(t'(\alpha, 0)) - x(s'(\alpha, 0)) \leq d(\alpha, 0)$. If $i = 1$, then $\alpha \in E_0$ and $d(\alpha) = 0$. Therefore, $0 \leq x(t(\alpha)) - x(s(\alpha)) \leq d(\alpha) = 0$ and $0 \leq x(s(\alpha)) - x(t(\alpha)) \leq 0 = d(\alpha, 1)$. Since $s'(\alpha, 1) = t(\alpha)$ and $t'(\alpha, 1) = s(\alpha), 0 \leq x(t'(\alpha, 1)) - x(s'(\alpha, 1)) \leq d(\alpha, 1)$. Therefore, $x \in C(T(G))$.

Conversely, if $x \in T(G)$ and $\alpha \in E$, then $0 \leq x(t'(\alpha, 0)) - x(s'(\alpha, 0)) \leq d(\alpha, 0)$. As before, $s'(\alpha, 0) = s(\alpha), t'(\alpha, 0) = t(\alpha)$, and $d'(\alpha, 0) = d(\alpha)$, so that $0 \leq x(t(\alpha)) - x(s(\alpha)) \leq d(\alpha)$ 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 $x_1 \preceq_G x_2$, $x_1(t(\alpha)) - x_1(s(\alpha)) \leq x_2(t(\alpha)) - x_2(s(\alpha))$, so that $x_1(t'(\alpha, 0)) - x_1(s'(\alpha, 0)) \leq x_2(t'(\alpha, 0)) - x_2(s'(\alpha, 0))$. If $i = 1$, then $\alpha \in E_0, x_i(t(\alpha)) = x_i(s(\alpha))$, and

$$x_1(t'(\alpha, 1)) - x_1(s'(\alpha, 1)) = 0 \leq 0 = x_2(t'(\alpha, 1)) - x_2(s'(\alpha, 1)).$$

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 $x_1 \preceq_{T(G)} x_2, x_1(t'(\alpha, 0)) - x_1(s'(\alpha, 0)) \leq x_2(t'(\alpha, 0)) - x_2(s'(\alpha, 0))$. Therefore, $x_1(t(\alpha)) - x_1(s(\alpha)) \leq x_2(t(\alpha)) - x_2(s(\alpha))$, so that $x_1 \preceq_G x_2$. $\square$

Given a a weighted, directed graph, $G = (V, E, d, s, t)$, let $[v] = \{v' \in V \mid v' \sim v\}$. We wish to define a graph $P(G) = (V', E', d', s', t')$, 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 $E_{s',t'} \equiv \{\alpha \in E \mid s(\alpha) \in s', e(\alpha) \in t'\} \neq \emptyset$, 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 $t'(\alpha) = [t(\alpha)]$. 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 9.6. *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, $p\left(\max C(P(G))\right) = \max C(G)$.*

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 $0 \leq x(t(\alpha)) - x(s(\alpha)) \leq d(\alpha)$. By definition, $x(t(\alpha)) - x(s(\alpha)) = x'([t(\alpha)]) - x'([s(\alpha)])$. If $s(\alpha) \sim t(\alpha)$, this is 0 and we are done. Otherwise, letting $t' \equiv [t(\alpha)]$ and $s' \equiv [s(\alpha)]$, we have $s(\alpha) \prec_G t(\alpha)$ and $\alpha \in E_{s',t'}$, so that $x'([t(\alpha)]) - x'([s(\alpha)]) = x'(t') - x'(s') = x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right)$. Since $x' \in C(P(G))$, $0 \leq x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right) \leq d\left(\alpha_{s',t'}\right) \leq d(\alpha)$. Since $x(t(\alpha)) - x(s(\alpha)) = x'\left(e'\left(\alpha_{s',t'}\right)\right) - x'\left(s'\left(\alpha_{s',t'}\right)\right)$, we have shown that $x \in C(G)$.

Now observe that, by Theorem 9.4, any $x \in C(G)$ is constant on equivalence classes, $[v]$, so that the inverse mapping, $p^{-1}(x)([v]) = x(v)$, is well-defined. If $x' = p^{-1}(x)$, we must show that $x' \in C(P(G))$. If $\alpha \in E'$, then $\alpha = \alpha_{s',t'}$ with $s'(\alpha) = s'$ and $t'(\alpha) = t'$. Moreover, $s' = [s(\alpha)]$ and $t' = [t(\alpha)]$ and $x'(e'(\alpha)) - x'(s'(\alpha)) = x'([t(\alpha)]) - x'([s(\alpha)]) = x(t(\alpha)) - x(s(\alpha))$, 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 $t = t(\alpha)$, $s = s(\alpha)$, and observe that $p(x_i)(t) - p(x_i)(s) = x_i([t]) - x_i([s])$. 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 $p(x_1)(t) - p(x_1)(s) = 0 \leq 0 = p(x_1)(t) - p(x_1)(s)$. 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, $p\,(x_1)\,(t) - p\,(x_1)\,(s) \le p\,(x_1)\,(t) - p\,(x_1)\,(s)$. Therefore, $p\,(x_1) \le p\,(x_2)$.

Conversely, assume that $x_1' \preceq x_2'$ with $x_i' \in C(G)$. Let $x_i' = p\,(x_i)$ and consider $\alpha \in E'$. Then $\alpha = \alpha_{[s],[t]}$, with $s(\alpha) = s$ and $t(\alpha) = t$. Since $s'(\alpha) = [s]$ and $t'(\alpha) = [t]$, we must show that $x_1([t]) - x_1([s]) \le x_2([t]) - x_2([s])$. Since $x_1' \preceq x_2'$, $x_1'(t) - x_1'(s) \le x_2'(t) - x_2'(s)$. By definition, $x_i'(t) = x_i([t])$ and $x_i'(s) = x_i([s])$. Therefore, $x_1([t]) - x_1([s]) \le x_2([t]) - x_2([s])$, as desired, so that $x_1 \preceq x_2$. $\square$

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 $x\,(t(\alpha)) - x\,(s(\alpha)) = d(\alpha)$. Specifically, given a weighted, directed graph without self-loops, $G = (V, E, d, s, t)$, we will say that an edge, $\alpha \in E$, is *feasible* iff $d(\alpha) = \min_{t(\beta)=t(\alpha)} d(\beta)$. 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 $R_\alpha(G) = (V, E, d', s, t)$ with

$$d'(\beta) \equiv \begin{cases} d(\beta) - d(\alpha), & \text{if } t(\beta) = t(\alpha) \\ d(\beta) + d(\alpha), & \text{if } s(\beta) = t(\alpha) \\ d(\beta), & \text{otherwise} \end{cases} .$$

As before, there is a corresponding mapping, r_α, taking valuations on $R_\alpha(G)$ to valuations on G.

Theorem 9.7. *Consider the mapping r_α, given by*

$$r_\alpha\,(x')\,(v) \equiv \begin{cases} x'(v) + d(\alpha), & \text{if } v = t(\alpha) \\ x'(v), & \text{otherwise} \end{cases}$$

This is bijective, order-preserving correspondence between:

$$\text{domain}\, r_\alpha \equiv \left\{ x' \in C(R(G)) \mid x'(t(\beta)) - x'(s(\beta)) \ge d(\alpha),\ \textit{if}\, t(\alpha) = s(\beta) \right\}$$

and

$$\text{im}\, r_\alpha = \left\{ x \in C(G) \mid x(t(\beta)) - x(s(\beta)) \ge d(\alpha),\ \textit{if}\, t(\alpha) = t(\beta) \right\} .$$

Proof. We first show that

$$x \equiv r_\alpha\,(x') \in C' \equiv \left\{ x \in C(G) \mid x(t(\beta)) - x(s(\beta)) \ge d(\alpha),\ \text{if } t(\alpha) = t(\beta) \right\},$$

for any $x' \in \operatorname{domain} r_\alpha$. Given $\beta \in E$, by assumption, $0 \leq x'(e(\beta)) - x'(s(\beta)) \leq d'(\beta)$, with $d(\alpha) \leq x'(e(\beta)) - x'(s(\beta))$, if $t(\alpha) = s(\beta)$. If $t(\beta) = t(\alpha)$, then $d'(\beta) = d(\beta) - d(\alpha)$, $x'(t(\beta)) = x(t(\beta)) - d(\alpha)$, and $x'(s(\beta)) = x(s(\beta))$, since $s(\beta) \neq t(\beta) = t(\alpha)$. Therefore, $0 \leq d(\alpha) \leq x(t(\beta)) - x(s(\beta)) \leq d(\beta)$. In particular, $x(t(\alpha)) - x(s(\alpha)) = d(\alpha)$.

Similarly, if $s(\beta) = t(\alpha)$, then $d'(\beta) = d(\beta) + d(\alpha)$, $x'(s(\beta)) = x(s(\beta)) - d(\alpha)$, and $x'(t(\beta)) = x(t(\beta))$, since $t(\beta) \neq s(\beta) = t(\alpha)$. Since $d(\alpha) \leq x'(e(\beta)) - x'(s(\beta)) \leq d'(\beta)$, $0 \leq x(t(\beta)) - x(s(\beta)) \leq d(\beta)$.

Finally, if $t(\alpha) \neq s(\beta), t(\beta)$, then $d'(\beta) = d(\beta)$, $x'(s(\beta)) = x(s(\beta))$, and $x'(t(\beta)) = x(t(\beta))$. Thus, $0 \leq x(t(\beta)) - x(s(\beta)) \leq d(\beta)$ for all $\beta \in E$, with $d(\alpha) \leq x(t(\beta)) - x(s(\beta))$, when $t(\alpha) = t(\beta)$. 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 $v = t(\alpha)$, 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 $x \equiv r_\alpha(x')$, then for any $s, t \in V$, notice that $x_i(t) - x_i(s) = r_\alpha(x_i)(t) - r_\alpha(x_i)(s) - c$, where $c = -d(\alpha), d(\alpha)$, or 0, depending on whether s and/or t equal $t(\alpha)$. In any case, $x_1(t) - x_1(s) \leq x_2(t) - x_2(s)$ iff $r_\alpha(x_1)(t) - r_\alpha(x_1)(s) \leq r_\alpha(x_2)(t) - r_\alpha(x_2)(s)$, for all $s, t \in V$. In particular, $x_1 \preceq x_2$ iff $r_\alpha(x_1) \preceq r_\alpha(x_2)$. $\square$

Lemma 9.8. *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 $\beta \in E$, $x(t(\beta)) - x(s(\beta)) \geq x'(t(\beta)) - x'(s(\beta))$. By Theorem 9.7, if $t(\alpha) = t(\beta)$, $x'(t(\beta)) - x'(s(\beta)) \geq d(\alpha)$. Therefore, $x(t(\beta)) - x(s(\beta)) \geq d(\alpha)$, and again by Theorem 9.7, $x \in \operatorname{im} r_\alpha$. $\square$

Lemma 9.9. $r_\alpha(\max C(R_\alpha(G)) \cap \operatorname{domain} r_\alpha) \subset \max C(G)$.

Proof. Let $x_1 \in \max C(R_\alpha(G)) \cap \operatorname{domain} r_\alpha$ and $r_\alpha(x_1) \preceq x \in C(G)$. By Lemma 9.8, $x = r_\alpha(x_2)$, for $x_2 \in \operatorname{domain} r_\alpha$. Therefore, by Theorem 9.6, $x_1 \preceq x_2$, which implies that $x_1 \sim x_2$, since a_1 is maximal. Thus, $r_\alpha(x_1) \sim r_\alpha(x_2) = x$, again by Theorem 9.6. In particular, $r_\alpha(x_1) \in \max C(G)$. $\square$

Combining these results gives Algorithm 6 which computes Pareto optimal rankings.

Theorem 9.10. *Algorithm 6 produces a solution to* ***Rank***(G).

Algorithm 6 To Solve **Rank**(G).

1: **function** $x = \mathrm{rk}\,(G)$ {
2:      $G' = P(T(G))$
3:      **if** $\left(G'.\mathrm{numVertices} == 1\right)$
4:          **return**$(p(0))$;
5:      Choose a feasible $\alpha \in E$ such that $t(\alpha)$ is maximal.
6:      $G' = R_\alpha\left(G'\right)$
7:      **return**$\left(r_\alpha\left(p\left(\mathrm{rk}\left(G'\right)\right)\right)\right)$;
8: }

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 9.2, the constant ranking $x(v) = 0$ is clearly the unique feasible ranking, up to equivalence. Moreover, Theorems 9.6 and 9.5 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 $t(\alpha) = s(\beta)$. In particular, $\mathrm{domain}\,r_\alpha = C\left(R\left(G\right)\right)$, so that we may apply Lemma 9.9, along with Theorems 9.6 and 9.5 to prove that line 7 is correct. $\square$

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.

9.3   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.

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. One approach that Ailon (2007) use is to define a Markov process and use it to determine an aggregate ranking. We conjecture that we may similarly define a GSA, so that its associated energy function yields a meaningful aggregate ranking. In fact, we conjecture that a minor variant of Algorithm 4 is essentially the same as Algorithm 6, so that the Pareto optimal ranking it computes may

also be interpreted in terms of a Markov process.

Chapter 10

QuickRank

10.1 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 chapter, 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 chapter, 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 10.1, 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).
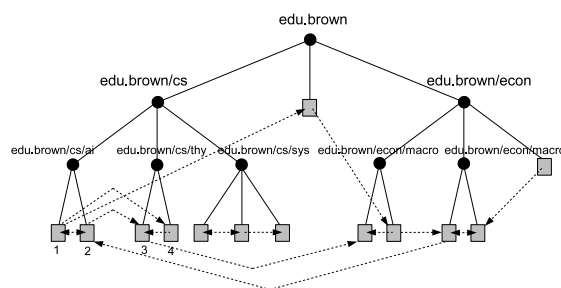

Figure 10.1: A sample hierarchical social network.

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 10.1. 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 chapter purports to contribute to the literature on social network analysis by introducing the QuickRank algorithm. As suggested by the previous example, Quick-Rank 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 10.3, 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 10.4, 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 10.8. A discussion of related work is deferred to the QuickRank technical report, currently in preparation.

10.2   A Unified View of Flat Ranking Algorithms

QuickRank is parameterized by a flat (i.e., non-hierarchical) ranking algorithm, or a "Base-Rank" 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.

10.2.1   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 Base-Rank 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.

10.2.2   Sample BaseRank Procedures

We now describe how the four ranking schemes mentioned in the introduction (i.e., in-degree, outdegree, eigenvector centrality and PageRank) can be viewed BaseRank proce-dures. 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 10.2).

The indegree and outdegree of individual i are defined respectively, as follows: given an initial judgment graph R,

$$\text{IN}(i) = \sum_j R_{ij} \qquad \text{OUT}(i) = \sum_j R_{ji} \tag{10.1}$$

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 $M_\epsilon = (1 - \epsilon)M + \epsilon vJ^t$. 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 10.1. *If M is a Markov matrix and $M_\epsilon = (1 - \epsilon)M + \epsilon vJ^t$, then*

$$v_\infty = \lim_{k \to \infty} M_\epsilon^k v_0 = \epsilon \sum_{i=0}^\infty (1 - \epsilon)^i M^i v \tag{10.2}$$

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 10.2. *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, $R_\alpha = vJ^t$.*

10.2.3 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 10.1 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 10.1, 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 from Theorem 10.2 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.

10.2.4 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 10.3. *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 10.3 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 10.1 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.

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 10.1: Some properties of ranking schemes.

Table 10.1 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 $\epsilon \sum_{i=0}^{\infty} (1 - \epsilon)^i M^{i+1}$, 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 10.4, we note that QuickRank preserves the spam-resistance of its BaseRank procedure, and we illustrate its potential to resist spam even further.

10.3   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

---

[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.

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 7) that is top-down and recursive, like many algorithms that operate on trees, the simplest way to visualize the 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 7 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 $Q = \begin{bmatrix} r^1 & \cdots & r^m \end{bmatrix}$, 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 7 QuickRank(node n)

1: **if** $(n.\text{isLeaf}())$
2:     **return**$(\langle n.\text{getJudgment}(), e_n \rangle)$;
3: $m = n.\text{numChildren}()$
4: **for** $(j = 1 \text{ to } m)$ {
5:     $\langle l^j, r^j \rangle \leftarrow \text{QuickRank}(n.\text{getChild}(j))$
6:     **for** $(k = 1 \text{ to } m)$ {
7:         $M_{kj} = \text{Sum}(l^j, n.\text{getChild}(k))$
8:     }
9: }
10: $P = \begin{bmatrix} l^1 & \cdots & l^m \end{bmatrix}$
11: $Q = \begin{bmatrix} r^1 & \cdots & r^m \end{bmatrix}$
12: $r = \text{BaseRank}(M, n.\text{getLocalPriorRanking}())$
13: **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.

10.4   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 10.2a. 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 10.2. The values in the posterior distributions have been rounded; hence, the ordinal rankings more precisely reflect the exact values in those distributions.
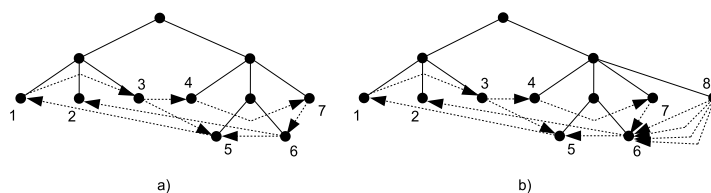

Figure 10.2: Two examples of hierarchical social networks.

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.

Table 10.2: BaseRanks and QuickRanks from Figure 10.2a 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

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, taking $\frac{1}{12}\{2, 2, 2, 2, 1, 1, 2\}$ as prior ranking and 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 Base-Rank procedures that do not exploit hierarchies. To demonstrate this phenomenon, in Figure 10.2b, 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 10.3: Fig. 10.2b 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 10.3. 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 10.3 shows that indegree produces the same rankings as in Table 10.2, that is, *without* the sybil, whether we exploit the hierarchy or not. In general, Theorem 10.3 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.

10.5   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).

---

[6]http://trec.nist.gov/

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.

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 10.2.2. We used the URL hierarchy as described in section 10.2.2, 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

---

[7]http://ir.dcs.gla.ac.uk/test_collections/

[8]http://lucene.apache.org/java/docs/

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*: $AP = \sum_{n=1}^{|H|} P@n \frac{\left[H_n - H_{n-1} \subset Qrels\right]}{|Qrels|}$

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 10.4-10.6. 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.

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 authority, we ran our experiments for various values of $alpha$ to discover a proper value, which ranged from .95 to .99 over the three years.

---

[9]While $|H_n| = n$, we write it out to show the symmetry in the definitions of precision and recall.

Table 10.4: 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	-

Table 10.5: 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	-

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 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 10.4-10.6, 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 10.4, 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 10.6: 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	-

10.6 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.

10.7 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.

10.8 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:

---

[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.

> 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 Quick-Rank, 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 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.

---

[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.

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 $\{v_i \in V \mid 1 \le i \le k\}$ is said to be *linearly independent* iff $\alpha_i = 0$, for all $1 \le i \le k$, 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.

Alain Trouvé. Cycle decompositions and simulated annealing. *SIAM J. Control Optim.*, 34:966–986, 1996.

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, MA, 1996.

H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, January 1993.