Sampling Spanning Trees
Theory and Algorithms

Nathan Benedetto Proença
Supervisor: Marcel K. de Carli Silva

July 19, 2019

Contents

Chapter 0

Introduction

This is a work in mathematics. It has no experiments, no data, and almost no pseudocode. It does have a myriad of propositions, lemmas and theorems. Yet, its unifying feature is an algorithmic problem. To have one applied question as a goal is often a rewarding setting. The birth of calculus was related with the problem of describing the laws governing the motion of celestial bodies, and it is but one example in which a problem in physics resulted in mathematical theory. However, one must acknowledge that an interesting problem may not fit under a single theory. The ever present human need to label ideas cannot possibly capture the nuances present in every interesting question. This applies to the labels Mathematics and Computer Science just as much as it applies to sublabels like Graph Theory and Linear Algebra. An investigation cannot restrict itself to a given domain of knowledge; it must unreservedly inspect whichever ideas are necessary to reach the aspired goal. This is the approach in the current study. It does not pursue a theoretical breakthrough — even though it does have new developments — but it aims at solidifying the foundations on which the results are based.

In general, a connected graph has an exponential number of spanning trees. Some applications have an interest in attaching a probability distribution to this set, and then to sample a spanning tree accordingly. Two compelling applications that need to do this kind of sampling in polynomial time can be found in [9] and [3] . In the first, Frieze et al. describes an algorithm to generate expander graphs, which have a rich set of applications, as can be seen in [13]. In the second, Asadpour et al. give an approximation algorithm for the asymmetric traveling salesman problem.

The problem itself demands the work to be broad. Moreover, since such problem is not only solved once, but twice, with mostly disjoint theoretical machinery, there is a gigantic amount of definitions to be made and propositions to be proven, from many distinct areas. Unsurprisingly, probability and measure theory play an important role. It is not only in the formulation of the problem, however. Chapter 3 defines a random walk rigorously, and establishes a precise language in which a couple of useful ideas can be inspected. Those ideas are at the core of the second algorithm we present, first developed by Aldous [1] and Broder [5], which is described in Chapter 4. Linear algebra, via algebraic combinatorics, is the setting for Chapter 2. It describes an algorithm, first developed by Kulkarni in [18], whose essence is a result by Kirchhoff [17]. Furthermore, the phrasing and language used in the text is influenced by the computer science background of the author. The elegance and expressiveness found in functional programming and category theory make themselves present, mostly on the prominence of function composition.

In broad terms, this monograph can be described as an attempt to make meaningful ideas, which are related to a specific problem, precise. Rigour is seen as the technique of the mathematician. Not a crutch, but a tool. This attitude pays off in a beautiful new development, found in Section 2.1. There is scarce literature properly defining matrices and common operations in it. This lack of language make unprecise some of the concepts needed to state and prove the Cauchy-Binet Formula. The solution found, namely, function matrices, reduce the absent definitions to function composition and simpler matrix operations. Furthermore, it even gives new insight into matrices defined in algebraic graph theory, relating them in a cleaner way to the formal definition of a graph.

Chapter 1

Preliminaries

1.1 Basic Definitions and Results

We work under the assumption that the reader is familiar with basic set theory concepts. We begin by describing some notations that will be heavily used in the work. Let A and B be sets. We denote by

$$B^A = \{\, f : f \text{ is a function with domain } A, \text{ taking values on } B\}.$$

We also assume that a function $f\colon A \to B$ has an inverse if and only if it is injective and surjective. Such inverse will be denoted as f^{-1}. Moreover, for a given set A we denote by $\mathcal{P}(A)$ the set of subsets of A. The set $\mathcal{P}(A)$ is the *powerset of* A. We denote by $A \setminus B$ the set of elements which belong to A but does not belong to B. Moreover, we define *the preimage with respect to* f as the function $f^{-1}\colon \mathcal{P}(B) \to \mathcal{P}(A)$ defined by

$$f^{-1}(S) = \{\, a \in A : f(a) \in S\},$$

for every $S \in \mathcal{P}(B)$.

Proposition 1.1. Let A and B be sets. Let $f\colon A \to B$ be a function. Then

1. For every subset $S \subseteq B$, we have that $f^{-1}(B \setminus S) = A \setminus f^{-1}(S)$.
2. For every collection $\mathcal{B} \subseteq \mathcal{P}(B)$, we have that $f^{-1}(\bigcup \mathcal{B}) = \bigcup_{S \in \mathcal{B}} f^{-1}(S)$.

Let A and B are sets, and $f\colon A \to B$ is a function. For every subset $S \subseteq A$, we denote by

$$f{\restriction_S}\colon S \to B$$

the function obtained from restricting the domain of A to S. Function composition will be so important in the current text that we reserve the most important of notations for it. Let A, B and C be sets, and let $f\colon B \to C$ and $g\colon A \to B$. We define the function $fg\colon A \to C$ as

$$fg(a) = f(g(a)),$$

for every $a \in A$. Finally, we denote by $\mathbb{N}$ the set of natural numbers, and for every $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1, \ldots, n\}$. The set of real numbers is denote by $\mathbb{R}$.

We now focus on some concepts that are less common. The reader is invited to take a look at Equation (3.4), which is used to define a random walk on a graph, as a use case for the notations being introduced.

The first one is the *Iverson bracket*. Let P be any predicate, i.e., true or false expression. Then

$$[P] \coloneqq \begin{cases} 1, & \text{if } P \text{ is true,} \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, if a given predicate P is false, the whole expression containing $[P]$ "short-circuits" to zero. This is helpful since $[P]$ can multiply an expression which is not defined if P is nonzero. In Equation (3.4), the function w is only defined on the set A; hence, if $ij \in V \times V$ is not in A, the expression $w(ij)$ is meaningless.

We also work with *multisets*. A multiset is a pair (S, c), where S is a finite set and $c \colon S \to \mathbb{N}$ is a function. For a given $s \in S$, the number $c(s)$ is the *multiplicity* of s, and represents "how many copies of s are in S". Accordingly, if $S \subseteq \mathbb{R}$, we define

$$\begin{aligned} \sum S &\coloneqq \sum_{s \in S} c(s)s, \\ \prod S &\coloneqq \prod_{s \in S} s^{c(s)}. \end{aligned}$$

A finite set S can be promoted into a multiset with the function $c(s) = 1$ for every $s \in S$. Usually a multiset is denoted as S, whenever the multiplicity can be inferred from the context. Moreover, if (S, c) is a multiset, and $f \colon S \to T$ is any function, we define a new multiset $(f(S), c')$ as

$$\begin{aligned} f(S) &\coloneqq \{\, f(s) : s \in S\}, \\ c'(y) &\coloneqq \sum_{s \in S} [f(s) = y] c(s) \quad \forall y \in f(S). \end{aligned}$$

Given a finite set S and a function $f \colon S \to \mathbb{R}$, we have that $f(S)$ is a multiset as defined above. It makes sense then to work with $\sum f(S)$. The denominator of Equation (3.4) is an example of this notation.

Definition 1.2. A relation $\leq$ in a set A is a *partial order* if

(i) the relation $\leq$ is *reflexive*, that is, for every $a \in A$ it holds that $a \leq a$,

(ii) the relation $\leq$ is *antisymmetric*, that is, for every $a, b \in A$, if $a \leq b$ and $b \leq a$ then $a = b$,

(iii) the relation $\leq$ is *transitive*, that is, for every $a, b \in A$, if $a \leq b$ and $b \leq c$, then $a \leq c$.

When $\leq$ is a partial order on A, we also say that $(A, \leq)$ is a *partially ordered set*.

Definition 1.3. A partially ordered set $(A, \leq)$ is *totally ordered* if the relation $\leq$ is *total*, that is, if for every $a, b \in A$ it holds that $a \leq b$ or $b \leq a$.

Partial and total orders will be commonplace throughout the work. When working with the real numbers, for example, the total order in it is crucial in defining the idea of convergence. In other places, it is merely a convenience, used mostly as a way to represent a "choice", as it is the case in Section 2.4.

Proposition 1.4. Let V be a finite set and let $\mu \in \mathbb{R}^V_+$ be such that $\sum_{i \in V} \mu_i = 1$. Then for every $f \in \mathbb{R}^V$, there exists $i \in V$ such that

$$\sum_{j \in V} \mu_j f_j \leq f_i.$$

Proof. We proceed to prove the statement by contradiction.

Suppose there exists $\mu \in \mathbb{R}^V_+$ such that $\sum_{i \in V} \mu_i = 1$ and $f \in \mathbb{R}^V$ such that for every $i \in V$ it holds that $f_i < \alpha$, where $\alpha \coloneqq \sum_{i \in V} \mu_i f_i$. Then

$$\alpha = \sum_{i \in V} \mu_i f_i < \sum_{i \in V} \mu_i \alpha = \alpha \sum_{i \in V} \mu_i = \alpha.$$

In other words, $\alpha < \alpha$, and the proof is done. □

1.2 Graph Theory

A *graph* is an ordered triple (V, E, ψ), where V and E are finite sets and $\psi \colon E \to \binom{V}{1} \cup \binom{V}{2}$. The elements of the set V are called *vertices*. The elements of the set E are called *edges*. The function ψ is called the *incidence function*. An edge $e \in E$ is said to be *incident* to the vertices that belong to $\psi(e)$. Two vertices i and j are said to be *adjacent*, or, likewise, j is said to be *adjacent* to i, if $\{i, j\} \in \psi(E)$. Moreover, the sets V and E are also denoted by $V(G)$ and $E(G)$, respectively.

A *digraph* is an ordered triple (V, A, ψ), where V and A are finite sets and $\psi\colon A \to V \times V$. The elements of the set V are called *vertices* and the function ψ is called *the incidence function*, just like before. However, the elements of the set A are called *arcs*, to emphasize their difference in nature from the edges of a graph. This difference renders the meaning of "adjacent" ambiguous, and its use will be avoided. For an arc $a \in A$, if $\psi(a) = (i, j)$, the vertex i is said to be the *tail* of the arc, the vertex j is said to be the *head* of the arc. Moreover, if $\psi(a) = (i, j)$, then a is said to be incident on i, and a is said to be incident on j. Moreover, the sets V and A are also denote by $V(D)$ and $A(D)$, respectively.

Let $G = (V, E, \psi)$ be a graph. For every $S \subseteq E$, we denote by $G[S]$ the graph $(V, S, \psi\restriction_S)$. Likewise, if $D = (V, A, \psi)$ is a digraph, for every $S \subseteq A$ we denote by $D[S]$ the digraph $(V, S, \psi\restriction_S)$.

There is now a need to pause and pay respect to tradition. There are some notations that are widespread for their simplicity, and should be properly explained according to the definitions.

First of all, even though arcs and edges are different, a notation that masks their differences is widely adopted. In both contexts, graphs and digraphs, ij will be used, and it is hoped the reader will notice and correctly parse it as (i, j) when it is an arc, and as $\{i, j\}$ when it is an edge.

Also, a graph is commonly thought of as a symmetric relation E on a finite set V. This interpretation usually casts aside the case of *parallel edges*, i.e., of distinct edges $e \in E$ and $f \in E$ such that $\psi(e) = \psi(f)$. To ignore such cases is precisely to require that the incidence function is injective. Note that, in such cases, the set $\psi(E)$ uniquely determines the graph. When this happens, the incidence function can be omitted, and the graph can be denoted as $G = (V, E)$, when what is actually meant is $G = (V, \psi(E), (x \mapsto x))$. The same reasoning should be applied when "a digraph $D = (V, A)$" is encountered within the text.

Another special case when working with graphs and digraphs are loops. Let $G = (V, E, \psi)$ be a graph. A *loop* is an edge $e \in E$ such that $\psi(e) = \{e\}$. Similarly, if $D = (V, A, \psi)$ is a digraph, a loop is an arc $a \in A$ such that $\psi(a) = (a, a)$. A graph or digraph with no parallel edges and no loops is a *simple* graph.

Let $D = (V, A, \psi)$ be a digraph. Let $\pi\colon V \times V \to \binom{V}{1} \cup \binom{V}{2}$ be defined by

$$\pi(i, j) \coloneqq \{i, j\}.$$

The *underlying graph* of D is the graph $G \coloneqq (V, A, \pi\psi)$. The function π encodes the idea of "forgetting" the orientation of an arc. For such a reason, another way of stating that D is a digraph and G is its underlying graph is to state that D is an *orientation* of G.

It is also quite convenient to associate more information with a graph. A *weighted graph* is a pair (G, w), where $G = (V, E, \psi)$ is a graph and $w\colon E \to \mathbb{R}$. We may also denote a weighted graph as $G = (V, E, \psi, w)$. Similarly, a *weighted digraph* is a pair (D, w), where $D = (V, A, \psi)$ is a digraph and $w\colon A \to \mathbb{R}$. We may also denoted it as $D = (V, A, \psi, w)$.

It is important to note that all the nomenclature about graphs and digraphs extends itself naturally to weighted graphs and digraphs. For example, given a weighted digraph (D, w), the weighted graph (G, w), with G being the underlying graph of D, will be called underlying weighted graph of (D, w).

Let $G = (V, E, \psi)$ be a graph. Let $\leq$ be a total order in V. The *symmetric digraph from* G is the digraph $D \coloneqq (V, A, \phi)$ defined as $A \coloneqq E \times 2$ and

$$\phi(ij, k) \coloneqq \begin{cases} \min\{i, j\}, & \text{if } k = 0, \\ \max\{i, j\}, & \text{if } k = 1. \end{cases}$$

It has an arc in both directions for each edge in the original graph. Moreover, if (G, e) is a weighted graph, we define $\hat{w} \in \mathbb{R}^A$ such that

$$\hat{w}(ij, 0) = \hat{w}(ij, 1) = w(ij).$$

In such a setting, $(D, \hat{w})$ is the symmetric weighted digraph from the weighted graph (G, w).

Let $D = (V, A, \psi)$ be a digraph. For every arc $a \in A$, the *arc contraction of* a is the digraph $D/a \coloneqq (f(V), A \setminus \{a\}, \hat{f}\psi)$, where $f\colon V \to V \cup \{a_0\}$ is defined as

$$f(k) \coloneqq \left\{ \begin{array}{ll} a_0, & \text{if } \exists j \in V \text{ such that } \psi(a_0) \in \{jk, kj\} \\ k, & \text{otherwise} \end{array} \right. ,$$

and $\hat{f}\colon V \times V \to (V \cup \{a_0\}) \times (V \cup \{a_0\})$ is defined by $(i, j) \mapsto (f(i), f(j))$.

A *walk* in a graph G is a finite alternating sequence $(u_0, e_1, u_1, \ldots, e_\ell, u_\ell)$ of vertices and edges such that, for every $0 < i \leq \ell$,

$$\psi(e_i) = \{u_{i-1}, u_i\}.$$

This walk is from u_0 to u_ℓ. If we have that $i = u_0$ and $j = u_\ell$, then the walk is an ij-walk. The integer ℓ is called the *length* of the walk. Note that ℓ is precisely the number of edges in it, which is one more than the number of vertices. Similarly, a walk in a digraph D is a finite alternating sequence of vertices and arcs $(u_0, a_1, u_1, \ldots, a_\ell, u_\ell)$ such that, for every $0 < i \leq \ell$,

$$\psi(a_i) = (u_{i-1}, u_i).$$

Further down the road, the text will talk about "random walks". Beware: despite the name, a random walk on a graph is not a walk as defined above. It is actually a much more interesting mathematical object, that is connected to walks, but which will demand its own definition and machinery to be dealt with.

There are some concepts related to walks. A *trail* in a graph G is a walk $(u_0, e_1, \ldots, e_\ell, u_\ell)$ in G such that the map $i \mapsto e_i$, defined on $\{1, \ldots, \ell\}$, is injective. In other words, a trail is a walk where no edge appears twice. A *path* in a graph G is a walk $(u_0, e_1, \ldots, e_\ell, u_\ell)$ in G such that the map $i \mapsto u_i$, defined on $\{0, \ldots, \ell + 1\}$ is injective. In other words, a path is a walk where no vertice appears twice. Finally, a *cycle* in a graph G is a trail $(u_0, e_0, \ldots, e_\ell, u_\ell)$ in G such that $u_0 = u_\ell$. A graph that has no cycles is *acyclic.* Note that a cycle can have repeated vertices, but no repeated edges. The same terminology applies to digraphs, in the sense that trail, path, cycle and acyclic can be similarly defined in a digraph.

Two distinct vertices i and j in a graph are said to be *connected* if there exists a walk $(i, e_0, \ldots, e_\ell, j)$. A graph is said to be connected if every pair of vertices is connected.

A *subgraph* of a graph $G = (V, E, \psi)$ is a graph $H = (S, F, \phi)$, with $S \subseteq V$, $F \subseteq E$ and ϕ being the restriction of ψ on F. Likewise, a *subdigraph* of a digraph $D = (V, A, \psi)$ is a digraph $C = (S, B, \phi)$, with $S \subseteq V$, $B \subseteq A$ and ϕ being the restriction of ψ on B. The set of subgraphs of a graph G, when equipped with the relation "is a subgraph of", forms a lattice. This observation gives meaning to statements like *minimal subgraph* and *maximal subgraph.* The same idea applies to the set of subdigraphs of a digraph. A graph $H = (S, F, \psi)$ of $G = (V, E, \psi)$ is *spanning* if it is a subgraph such that $S = V$. Let $G = (V, E, \psi)$ be a graph. Let $S \subseteq V$. We denote by $G[S]$ the subgraph $(V, F, \psi\restriction_F)$, where

$$F \coloneqq \{\, e \in E : \psi(e) \subseteq S\}.$$

A *component* of a graph G is a maximal connected subgraph. It is interesting to note that "is connected to" defines a equivalence relation in the vertices of a graph, and a component is a equivalence class in it. Note, then, that a connected graph has only one component.

Definition 1.5. A *tree* is a connected acyclic graph.

A *spanning tree* of a graph is a spanning subgraph that is a tree. This is equivalent to say that it is a minimal connected spanning subgraph, i.e., a subgraph such that every spanning subgraph of it is not connected. The collection of sets of edges F such that (V, F) is a spanning tree is denoted as $\mathcal{T}_G$.

Let $G = (V, E, \psi)$ be a graph. We denote by δ the function $\delta \colon V \to \mathcal{P}(E)$ defined by

$$\delta(i) \coloneqq \{\, e \in E : \{i\} \subsetneq \psi(e)\},$$

for every $i \in V$. The integer $|\delta(i)|$ is called the *degree* of the vertex i. When working with digraphs, two such functions are defined. Let $D = (V, A, \psi)$ be a digraph. We denote by δ^{in} the function $\delta^{\text{in}} \colon V \to \mathcal{P}(A)$ defined by

$$\delta^{\text{in}}(i) \coloneqq \{\, e \in E : \exists j \in V \quad \psi(e) = ji\},$$

for every $i \in V$. The integer $\left|\delta^{\text{in}}(i)\right|$ is called the *in-degree* of vertex i. Similarly, we denote by δ^{out} the function $\delta^{\text{out}} \colon V \to 2^A$ defined by

$$\delta^{\text{out}}(i) \coloneqq \{\, e \in E : \exists j \in V \quad \psi(e) = ij\}$$

for every $i \in V$. The integer $|\delta^{\text{out}}(i)|$ is called the *out-degree* of vertex i.

Definition 1.6. An *r-arborescence* is a digraph $D = (V, A, \psi)$ such that $r \in V$ and

1. its underlying graph G is a tree and
2. for every $i \in V$,
$$\left|\delta^{\text{in}}(i)\right| = [i \neq r].$$

Whenever r is specified, the graph D is said to be an r-arborescence. For a fixed $r \in V$, the collection of sets of arcs B such that $D[B]$ is a r-arborescence is denoted as $\mathcal{T}_D(r)$. The set of all arborescences of a graph is denoted as $\mathcal{T}_D$.

Vertices in a tree whose degree equals to 1 are called *leafs.* Leafs are extremely helpful in proofs by induction. They are so important that the following statement, that ensures the existance of leafs, will be used without mention when working with trees.

Proposition 1.7. Let V be a finite set, and let $f \in \mathbb{R}^V$. There exists $k \in V$ such that
$$f(k) \leq \frac{1}{|V|} \sum_{i \in V} f(i).$$

Proof. Define $\mu \in \mathbb{R}^V_+$ to be $|V|^{-1}$ for every vertex. Then Proposition 1.4 ensures that there exists $k \in V$ such that
$$-\sum_{i \in V} \frac{1}{|V|} f(i) = \sum_{i \in V} (-f_i)\mu_i \leq -f(k).$$
It is enough to multiply both sides by -1. □

Theorem 1.8. Let $T = (V, E)$ be a tree with $|V| \geq 2$. Then there are at least 2 vertices of T with degreee 1.

Proof. The proposition above will be the main tool on this proof. First, note that if $T = (V, A)$ is a tree,
$$\frac{1}{|V|} \sum_{i \in V} |\delta(i)| = \frac{2(|V| - 1)}{|V|} = 2 - \frac{2}{|V|}.$$
Therefore, there exists a vertex $k \in V$ such that $|\delta(k)| \leq 2 - \frac{2}{|V|}$. Since $|\delta(k)|$ must be an integer, we have that $|\delta(k)| \leq 1$. Also, since every tree is connected, the degree of every vertex is at least 1, so that $|\delta(k)| = 1$.

To produce the second vertex with degree 1, suffices to repeat the argument, without the already leaf k. Note that
$$\frac{1}{|V| - 1} \sum_{i \in V \setminus \{k\}} |\delta(i)| = \frac{2(|V| - 1) - 1}{|V| - 1} = 2 - \frac{1}{|V| - 1}.$$

Therefore, there is a $j \in V$ with $|\delta(j)| \leq 1$, and as before, this implies that $|\delta(j)| = 1$, finishing the proof. □

Theorem 1.9. Let $D = (V, A, \psi)$ be digraph. Let $i \in V$. Suppose D is an i-arborescence, with $|V| \geq 2$. Then there exists $j \in V \setminus \{i\}$ with outdegree 1.

Proof. The underlying graph of D has at least two leafs, at least one of which is different from i. Let j be it. Then, since $j \neq i$, we have that $|\delta(j)| = 1$. Therefore, since
$$\left|\delta^{in}(j)\right| + \left|\delta^{out}(j)\right| = 1,$$
it follows that $\delta^{out}(j) = \varnothing$. □

1.3 Matrices and Determinant

We hope that the reader is familiar with the concept of a vector space. For every finite set U, we treat $\mathbb{R}^U$ as a vector space under the familiar operations.

Definition 1.10. Let U and V be finite sets. A *matrix* is a function $A\colon V \times U \to \mathbb{R}$. The elements of V are the *row indices* of A, and the elements of U are the *column indices* of A.

Definition 1.11. Let U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$. The *tranpose* A^T is a matrix in $\mathbb{R}^{U \times V}$ defined by

$$(i, j) \mapsto A_{ji}.$$

For a finite set U, a matrix $A \in \mathbb{R}^{U \times U}$ is *symmetric* if $A^\mathsf{T} = A$.

Definition 1.12. Let U, V, and T be finite sets. Let $A \in \mathbb{R}^{V \times U}$ and $B \in \mathbb{R}^{U \times T}$. The *product* AB is the matrix $AB\colon V \times T \to \mathbb{R}$ given by

$$(i, j) \mapsto \sum_{k \in U} A_{ik} B_{kj}.$$

The sets $\mathbb{R}^U$ and $\mathbb{R}^{U \times 1}$ will be used interchangeably. Such abuse is both possible and helpful. It is possible since there is a canonical isomorphism between such sets, and it is helpful since it reduces every matrix-vector product into a matrix-matrix product. Let $x \in \mathbb{R}^{U \times 1}$, we have that $x^\mathsf{T} \in \mathbb{R}^{1 \times U}$. Hence, for every $y \in \mathbb{R}^{U \times 1}$,

$$x^\mathsf{T} y = \sum_{i \in U} x_i y_i.$$

Proposition 1.13. Let U be a finite set.

(i) For every $x, y, z \in \mathbb{R}^U$ and $\alpha \in \mathbb{R}$, we have that $(\alpha x + y)^\mathsf{T} z = \alpha x^\mathsf{T} z + y^\mathsf{T} z$.

(ii) for every $x, y \in \mathbb{R}^U$, we have that $x^\mathsf{T} y = y^\mathsf{T} x$.

(iii) for every $x \in \mathbb{R}^U$, we have that $x^\mathsf{T} x \geq 0$. Moreover, $x^\mathsf{T} x = 0$ implies that $x = 0$.

Let U be a finite set. Then for every $i \in U$, define $e_i \in \mathbb{R}^U$ as

$$e_i(j) \coloneqq [i = j],$$

for every $j \in V$. This basis is the *canonical basis of* $\mathbb{R}^U$. Moreover, we define $\mathbb{1} \in \mathbb{R}^U$ as

$$\mathbb{1} \coloneqq \sum_{i \in U} e_i.$$

Note that both e_i and $\mathbb{1}$ have no reference of the vector space in which they are defined. However, whenever they appear in a calculation, the appropriate space is clear, so this raises no problems. Moreover, let $w \in \mathbb{R}^U$. The matrix $\operatorname{Diag}(w) \in \mathbb{R}^{U \times U}$ is defined as

$$\operatorname{Diag}(w) = \sum_{i \in U} w(i) e_i e_i^\mathsf{T}.$$

If we use the fact that for every $i, j \in U$ we have that

$$e_i^\mathsf{T} e_j = [i = j],$$

we can conclude that $\mathbb{1}^\mathsf{T} \operatorname{Diag}(w) \mathbb{1} = w^\mathsf{T} \mathbb{1}$.

Let V be a vector space over the $\mathbb{R}$, and $S \subseteq V$ is a finite subset. We denote by

$$\operatorname{span}(S) = \left\{ \sum_{s \in S} x(s)\, s : x \in \mathbb{R}^S \right\}.$$

Moreover, we use that this is a subspace. It is actually quite interesting how similiar this construction is to the ideia of σ-algebra generated by a set, which will be dealt in depth in Section 1.9. We assume, however, the reader is familiar with the current concept. Moreover, let $S \subseteq \mathbb{R}^U$. The set

$$S^\perp = \{\, x \in \mathbb{R}^U : x^\mathsf{T} s = 0 \quad \forall s \in S\}$$

is the *orthogonal complement of* S. We will also use the fact that

$$\left(S^\perp\right)^\perp = \operatorname{span}(S),$$

so that whenever S is a linear subspace, we have that $S = \left(S^\perp\right)^\perp$.

Proposition 1.14. Let U be a finite set. Then $x \in \mathbb{R}^U$ is zero if and only if for every $y \in \mathbb{R}^U$,

$$x^\mathsf{T} y = 0.$$

Proof. If $x = 0$, the thesis clearly holds. Suppose then that $x \in \mathbb{R}^U$ is such that for every $y \in \mathbb{R}^U$ it holds that $x^\mathsf{T} y = 0$. In particular, $x^\mathsf{T} x = 0$. Proposition 1.13 then ensures that $x = 0$. □

Definition 1.15. Let U and V be finite sets, and let $A \in \mathbb{R}^{V \times U}$. Given sets $S \subseteq U$ and $T \subseteq V$, a *submatrix* $A[T, S]$ is the matrix obtained from A by restricting its domain from $V \times U$ to $T \times S$.

Proposition 1.16. Let U,V, and T be finite sets. Let $A \in \mathbb{R}^{V \times U}$ and $B \in \mathbb{R}^{U \times T}$. Let $R \subseteq T$ and $S \subseteq V$. Then

$$(AB)[S, R] = A[S, U]B[U, R].$$

Proof. This is precisely what the submatrix definition means, applied to the product of two matrices. □

Proposition 1.17. Let U and V be finite sets. Let $A, B \in \mathbb{R}^{V \times U}$. If $i \in V$ is such that either $A^\mathsf{T} e_i = 0$ or $B^\mathsf{T} e_i = 0$, then

$$A^\mathsf{T} B = A^\mathsf{T}[U, V \setminus \{i\}]B[V \setminus \{i\}, U].$$

Proof. It is enough to expand the definitions:

$$\begin{aligned} A^\mathsf{T} B &= A^\mathsf{T}[U, V \setminus \{i\}]B[V \setminus \{i\}, U] + A^\mathsf{T}[U, \{i\}]B[\{i\}, U] \\ &= A^\mathsf{T}[U, V \setminus \{i\}]B[V \setminus \{i\}, U] + A^\mathsf{T} e_i e_i^\mathsf{T} B \\ &= A^\mathsf{T}[U, V \setminus \{i\}]B[V \setminus \{i\}, U]. \end{aligned}$$ □

Definitions around group theory can be found on any introductory text. A good reference is the first chapter of [14]. We use the definition of group homomorphism, the fact that a group isomorphism is a bijective homomorphism whose inverse is an homomorphism, the fact that composition of homomorphism is again an homomorphism.

Definition 1.18. Let V be a finite set. The *symmetric group of* V, denote as $\operatorname{Sym}(V)$ is the group of permutations of the set V, with composition as product.

Let V be a finite set. Let $i, j \in V$. Then the *transposition of* i *and* j is the function $(ij) \in \operatorname{Sym}(V)$ that fixes every element in $V \setminus \{i, j\}$, and swaps i and j. We assume the well-known facts that every permutation $\sigma \in \operatorname{Sym}(V)$ can be decomposed as a product of transpositions, and that, despite the decomposition not being unique, every decomposition of σ has length of the same parity. Hence we define $\operatorname{sgn}\colon \operatorname{Sym}(V) \to \{\pm 1\}$ as

$$\operatorname{sgn}(\sigma) \coloneqq (-1)^{N(\sigma)},$$

where $N(\sigma)$ is the length of any decomposition in transpositions of σ. Halmos [11] and Conrad [6] are interesting references on why this definition works. Most importantly, [6] ensures that the following proposition holds.

Proposition 1.19. Let V be a finite set. The function $\operatorname{sgn}\colon \operatorname{Sym}(V) \to \{\pm 1\}$ is the only nonconstant group homomorphism from $\operatorname{Sym}(V)$ into $\{\pm 1\}$.

Lemma 1.20. Let U and V be finite sets. Let $f\colon U \to V$ and $g\colon V \to U$ be bijective functions. Then

$$\operatorname{sgn}(fg) = \operatorname{sgn}(gf).$$

Proof. During this proof, denote by sgn_U the sign function defined on $\operatorname{Sym}(U)$, and sgn_V the sign function defined on $\operatorname{Sym}(V)$.

Define the function $F\colon \operatorname{Sym}(U) \to \operatorname{Sym}(V)$ as

$$F(\sigma) \coloneqq f\sigma f^{-1}.$$

Note that for every $\sigma_0, \sigma_1 \in \operatorname{Sym}(U)$, we have that $F(\sigma_0\sigma_1) = F(\sigma_0)F(\sigma_1)$, so that F is a group homomorphism. Moreover, the function $\sigma \mapsto f^{-1}\sigma f$ is its inverse, and also is a group homomorphism. Therefore, F is a group isomorphism.

Since both sgn_V and F are group homomorphisms, we have that $\operatorname{sgn}_V F$ is a group homomorphism. Moreover, since F is surjective and sgn_V is nonconstant, we can then conclude that $\operatorname{sgn}_V F$ also is nonconstant. Hence, $\operatorname{sgn}_V F$ is a nonconstant group homomorphism from $\operatorname{Sym}(U)$ into $\{\pm 1\}$. However, sgn_U is the only such function. Therefore,

$$\underset{U}{\operatorname{sgn}} = \underset{V}{\operatorname{sgn}} F.$$

Since f and g are bijective, we have that $gf \in \operatorname{Sym}(U)$. Applying the just proven equation, we have that

$$\operatorname{sgn}(gf) = \operatorname{sgn}(f\,gf\,f^{-1}) = \operatorname{sgn}(fg). \qquad \square$$

Proposition 1.21. Let V be a finite set, and let $U \subseteq V$. For every $\sigma \in \operatorname{Sym}(U)$, define

$$\hat{\sigma}(k) \coloneqq \begin{cases} \sigma(k), & \text{if } k \in U, \\ k, & \text{otherwise.} \end{cases}$$

Then $\operatorname{sgn}_V(\hat{\sigma}) = \operatorname{sgn}_U(\sigma)$.

Proof. Since $U \subseteq V$, any decomposition of σ in transpositions is also a decomposition of $\hat{\sigma}$. Hence, $\operatorname{sgn}_U(\sigma) = \operatorname{sgn}_V(\hat{\sigma})$. □

It is supposed that the reader is already familiar with the notion of determinant. However, for the treatment required on this paper, it is necessary to take a longer look on the definitions and properties of determinants.

Definition 1.22. Let U be a finite set. The *determinant* of a matrix $A \in \mathbb{R}^{U \times U}$ is

$$\det(A) \coloneqq \sum_{\sigma \in \operatorname{Sym}(U)} \operatorname{sgn}(\sigma) \prod_{i \in U} A_{i,\sigma(i)}.$$

Theorem 1.23. Let U be a finite set, and let $A \in \mathbb{R}^{U \times U}$. Then

$$\det(A) = \det(A^{\mathsf{T}}).$$

Proof. By definition,

$$\det(A) = \sum_{\sigma \in \operatorname{Sym}(U)} \operatorname{sgn}(\sigma) \prod_{i \in U} A_{i,\sigma(i)}.$$

Since σ is invertible and $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$, then

$$\det(A) = \sum_{\sigma \in \operatorname{Sym}(U)} \operatorname{sgn}(\sigma^{-1}) \prod_{i \in U} A_{\sigma^{-1}(i),i}.$$

Moreover, by noting that the function $\sigma \mapsto \sigma^{-1}$ from $\operatorname{Sym}(U)$ to itself is bijective, one can change the summation range and obtain that

$$\det(A) = \sum_{\tau \in \operatorname{Sym}(U)} \operatorname{sgn}(\tau) \prod_{i \in U} A_{\tau(i),i} = \sum_{\tau \in \operatorname{Sym}(U)} \operatorname{sgn}(\tau) \prod_{i \in U} (A^{\mathsf{T}})_{i,\tau(i)} = \det(A^{\mathsf{T}}). \qquad \square$$

Proposition 1.24. Let U be a finite set. Let $i \in U$ and let $A \in \mathbb{R}^{U \times U}$ be such that $Ae_i = 0$. Then for every $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^U$,

$$\det(A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}}) = \alpha \det(A + x e_i^{\mathsf{T}}) + \beta \det(A + y e_i^{\mathsf{T}}).$$

Proof. Define the function $f \colon \mathrm{Sym}(U) \to \mathbb{R}$ as

$$f(\sigma) \coloneqq \prod_{j \in U \setminus \{i\}} A_{\sigma(j),j}.$$

Note that for every $\sigma \in \mathrm{Sym}(U)$, we have the convient fact that all of the following expressions are the same:

$$f(\sigma) = \prod_{j \in U \setminus \{i\}} (A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}})_{\sigma(j),j} = \prod_{j \in U \setminus \{i\}} (A + x e_i^{\mathsf{T}})_{\sigma(j),j} = \prod_{j \in U \setminus \{i\}} (A + y e_i^{\mathsf{T}})_{\sigma(j),j}.$$

Moreover, we have that

$$\begin{aligned}
\det(A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}}) &= \det((A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}})^{\mathsf{T}}) \\
&= \sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) \prod_{j \in U} (A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}})_{\sigma(j),j} \\
&= \sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A_{\sigma(i),i} + \alpha x_{\sigma(i)} + \beta y_{\sigma(i)}) \prod_{j \in U \setminus \{i\}} (A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}})_{\sigma(j),j} \\
&= \sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A_{\sigma(i),i} + \alpha x_{\sigma(i)} + \beta y_{\sigma(i)}) f(\sigma).
\end{aligned}$$

Since $Ae_i = 0$, we have that $A_{\sigma(i),i} = 0$ for every $\sigma \in \mathrm{Sym}(U)$. We explore this to finish the proof:

$$\begin{aligned}
\det(A + \alpha x e_i^{\mathsf{T}} + \beta y e_i^{\mathsf{T}}) &= \sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(\alpha x_{\sigma(i)} + \beta y_{\sigma(i)}) f(\sigma) \\
&= \alpha \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) x_{\sigma(i)} f(\sigma) \right) + \beta \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) y_{\sigma(i)} f(\sigma) \right) \\
&= \alpha \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A_{\sigma(i),i} + x_{\sigma(i)}) f(\sigma) \right) + \beta \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A_{\sigma(i),i} + y_{\sigma(i)}) f(\sigma) \right) \\
&= \alpha \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A + x e_i^{\mathsf{T}})_{\sigma(i),i} f(\sigma) \right) + \beta \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma)(A + y e_i^{\mathsf{T}})_{\sigma(i),i} f(\sigma) \right) \\
&= \alpha \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) \prod_{j \in U} (A + x e_i^{\mathsf{T}})_{\sigma(j),j} \right) + \beta \left(\sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) \prod_{j \in U} (A + y e_i^{\mathsf{T}})_{\sigma(j),j} \right) \\
&= \alpha \det(A + x e_i^{\mathsf{T}}) + \beta \det(A + y e_i^{\mathsf{T}}).
\end{aligned}$$

□

Proposition 1.25. Let U be a finite set, and let $A \in \mathbb{R}^{U \times U}$. Let $\sigma \in \mathrm{Sym}(U)$. Then

$$\det(AP_\sigma) = \mathrm{sgn}(\sigma) \det(A).$$

Proof.

$$\begin{aligned}
\det(AP_\sigma) &= \sum_{\pi \in \mathrm{Sym}(U)} \mathrm{sgn}(\pi) \prod_{i \in U} (AP_\sigma)_{i,\pi(i)} \\
&= \sum_{\pi \in \mathrm{Sym}(U)} \mathrm{sgn}(\pi) \prod_{i \in U} A_{i,\sigma\pi(i)}.
\end{aligned}$$

We can change the summation indices with the bijection $\pi \mapsto \sigma\pi$, and finish the proof:

$$\begin{aligned}\det(AP_\sigma) &= \sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\sigma^{-1}\tau)\prod_{i\in U} A_{i,\tau(i)}\\ &= \sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\sigma^{-1})\operatorname{sgn}(\tau)\prod_{i\in U} A_{i,\tau(i)}\\ &= \operatorname{sgn}(\sigma^{-1})\sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\tau)\prod_{i\in U} A_{i,\tau(i)}\\ &= \operatorname{sgn}(\sigma^{-1})\det(A)\\ &= \operatorname{sgn}(\sigma)\det(A).\end{aligned}$$ □

Proposition 1.26. Let U be a finite set. Let $A \in \mathbb{R}^{U\times U}$. If there exists i and j in U such that $Ae_i = Ae_j$, we have that $\det(A) = 0$.

Proof. Let $\sigma \in \operatorname{Sym}(U)$ be the transposition (ij). Since $Ae_i = Ae_j$, it follows that $A = AP_\sigma$. Moreover, $\operatorname{sgn}(\sigma) = -1$, so that Proposition 1.25 finishes the proof:

$$\det(A) = \det(AP_\sigma) = -\det(A).$$ □

Proposition 1.27. Let U be a finite set, and let $A \in \mathbb{R}^{U\times U}$. If the set $\{\, Ae_i : i \in U\}$ is linearly dependent, then $\det(A) = 0$.

Proof. If $\{\, Ae_i : i \in U\}$ is linearly dependent, then there is $k \in U$ and $\alpha \in \mathbb{R}^U$ such that

$$Ae_k = \sum_{i\in U\setminus\{k\}} \alpha_i Ae_i.$$

Applying Proposition 1.24 to the matrix $A - Ae_ke_k^{\mathsf{T}}$ ensures that

$$\det(A) = \det(A - Ae_ke_k^{\mathsf{T}} + Ae_ke_k^{\mathsf{T}}) = \det\left((A - Ae_ke_k^{\mathsf{T}}) + \sum_{i\in U\setminus\{k\}} \alpha_i Ae_i\right) = \sum_{i\in U\setminus\{k\}} \alpha_i \det(A - Ae_ke_k^{\mathsf{T}} + Ae_ie_k^{\mathsf{T}}).$$

But every term in the summation is zero, since for every $i \in U \setminus \{k\}$ we have that

$$(A - Ae_ke_k^{\mathsf{T}} + Ae_ie_k^{\mathsf{T}})e_k = Ae_k - Ae_k + Ae_i = Ae_i,$$

so that Proposition 1.26 ensures its determinant is zero. □

1.4 Projections and Direct Sum

Let U and V be finite sets, and let $A \in \mathbb{R}^{V\times U}$. The *operator defined by* A is the function $A\cdot\colon \mathbb{R}^U \to \mathbb{R}^V$ defined as

$$x \mapsto Ax,$$

where the RHS is the matrix product between A and x. Conversely, given any linear transformation $L\colon \mathbb{R}^U \to \mathbb{R}^V$, there exists a matrix $A \in \mathbb{R}^{V\times U}$ such that $X = A\cdot$. Suffices to define, for every $i \in V$ and $j \in U$,

$$A_{ij} \coloneqq (Xe_j)_i.$$

The definition depends on the basis used for the spaces involved. However, unless otherwise stated, both will be the canonical basis, so that to define a matrix is enough to define the operator of the desired matrix.

Definition 1.28. Let U and V be finite sets. Let $A \in \mathbb{R}^{V\times U}$. Then the *range of* A, denoted by $\operatorname{Im}(A)$, is defined as

$$\operatorname{Im}(A) \coloneqq \{\, Ax : x \in \mathbb{R}^U\}.$$

Likewise, the *nullspace of* A, denoted by $\operatorname{Null}(A)$, is defined as

$$\operatorname{Null}(A) \coloneqq \{\, x \in \mathbb{R}^U : Ax = 0\}.$$

Proposition 1.29. Let U, V and W be finite sets. Let $A \in \mathbb{R}^{W \times V}$ and $B \in \mathbb{R}^{V \times U}$. Then

(i) $\mathrm{Null}(B) \subseteq \mathrm{Null}(AB)$, and

(ii) $\mathrm{Im}(AB) \subseteq \mathrm{Im}(A)$.

Proof. Let $x \in \mathrm{Null}(B)$. Then $Bx = 0$, and $ABx = 0$. This proves (i). For (ii), let $y \in \mathrm{Im}(AB)$. Then there exists $x \in \mathbb{R}^U$ such that $y = ABx$. Therefore, $Bx \in \mathbb{R}^V$ is such that $A(Bx) = y$. Hence, $y \in \mathrm{Im}(A)$. □

Theorem 1.30. Let U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$. Then

(i) $\mathrm{Null}(A^\mathsf{T}) = \mathrm{Im}(A)^\perp$, and

(ii) $\mathrm{Im}(A^\mathsf{T}) = \mathrm{Null}(A)^\perp$.

Proof. Let $x \in \mathbb{R}^U$. Proposition 1.14 implies that $x \in \mathrm{Null}(A^\mathsf{T})$ if and only if for every $y \in \mathbb{R}^U$,

$$0 = y^\mathsf{T} A^\mathsf{T} x = (Ay)^\mathsf{T} x.$$

Hence, $x \in \mathrm{Null}(A^\mathsf{T})$ if and only if $x \in \mathrm{Im}(A)^\perp$. To prove (ii), note that (i) applied to A^T ensures that $\mathrm{Null}(A) = \mathrm{Im}(A^\mathsf{T})^\perp$. Hence

$$\mathrm{Im}(A^\mathsf{T}) = (\mathrm{Im}(A^\mathsf{T})^\perp)^\perp = \mathrm{Null}(A)^\perp.$$

□

Proposition 1.31. Let T, U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$ and $B \in \mathbb{R}^{U \times T}$. Then $AB = 0$ if and only if

$$\mathrm{Im}(B) \subseteq \mathrm{Null}(A).$$

Proof. Let $x \in \mathbb{R}^T$. Since $ABx = 0$, we have that $Bx \in \mathrm{Null}(A)$. Since this holds for every $x \in \mathbb{R}^T$, we have that $\mathrm{Im}(B) \subseteq \mathrm{Null}(A)$.

Suppose then that $\mathrm{Im}(B) \subseteq \mathrm{Null}(A)$. Therefore, for every $x \in \mathbb{R}^T$, we have that $Bx \in \mathrm{Null}(A)$. Hence, $ABx = 0$. Since this holds for every x, we have that $AB = 0$. □

Proposition 1.32. Let U be a finite set, and let $A \in \mathbb{R}^{U \times U}$. If there exists a nonzero $x \in \mathrm{Null}(A)$, then $\det(A) = 0$.

Proof. If there exists a nonzero $x \in \mathbb{R}^U$ such that $Ax = 0$, we have that

$$0 = Ax = A\left(\sum_{i \in U} e_i e_i^\mathsf{T}\right) x = \sum_{i \in U} A e_i e_i^\mathsf{T} x = \sum_{i \in U} x_i A e_i.$$

Hence, the set $\{\, Ae_i : i \in U\}$ is linearly dependent, and Proposition 1.27 finishes the proof. □

Corollary 1.33. Let U be a finite set, and let $A \in \mathbb{R}^{U \times U}$. If $\det(A) \neq 0$, then A is invertible.

Proof. Suppose $\det(A) \neq 0$. The contraposition of Proposition 1.32 implies that $\mathrm{Null}(A) = \{0\}$. Hence, A is injective. Likewise, since $\det(A^\mathsf{T}) = \det(A)$, we have that A^T is injective. Theorem 1.30 then implies that $\mathrm{Im}(A) = \mathrm{Null}(A^\mathsf{T})^\perp$, so that $\mathrm{Im}(A) = \mathbb{R}^U$. Hence, A is also surjective, so that it has an inverse. □

Definition 1.34. Let $S, T \subseteq \mathbb{R}^U$ be linear subspaces. If $S \cap T = \{0\}$, and $\mathbb{R}^U = S + T$, we say that $\mathbb{R}^U$ is the *direct sum* of S and T, and denote that by

$$\mathbb{R}^U = S \oplus T.$$

Note that if $\mathbb{R}^U = S \oplus T$, then for every $x \in \mathbb{R}^U$ we have a unique pair $(y, z) \in S \times T$ such that $x = y + z$. To see this, note that if there were two pairs, (y_0, z_0) and (y_1, z_1), whose sum is x, we could write $y_0 + z_0 = y_1 + z_1$ and conclude that

$$y_0 - y_1 = z_1 - z_0$$

The LHS is in S, and the RHS is in T, so that both sides must be zero.

Since the direct sum gives for every vector $x \in \mathbb{R}^U$ a unique element $y \in S$, this defines a function.

Definition 1.35. Let $S, T \subseteq \mathbb{R}^U$ be linear subspaces such that $\mathbb{R}^U = S \oplus T$. For every $x \in \mathbb{R}^U$, let $(y, z) \in S \times T$ be such that $x = y + z$. The *projector on S along T* is the matrix $P_{S,T} \in \mathbb{R}^{V \times U}$ such that

$$P_{S,T} \cdot x = y.$$

Note that if $\mathbb{R}^U = S \oplus T$, then

$$I = P_{S,T} + P_{T,S}.$$

Proposition 1.36. Let $P \in \mathbb{R}^{U \times U}$. Then $P^2 = P$ if and only if P is the projector on $\mathrm{Im}(P)$ along $\mathrm{Null}(P)$.

Proof. For any $P \in \mathbb{R}^{U \times U}$, the fact that $I = P + (I - P)$ ensures that $\mathbb{R}^V = \mathrm{Im}(P) + \mathrm{Im}(I - P)$. First we show that $P^2 = P$ if and only if $\mathrm{Im}(P) \cap \mathrm{Im}(I - P) = \{0\}$.

Suppose $P^2 = P$. Note that this implies that

$$P(I - P) = (I - P)P = 0.$$

Let $x \in \mathrm{Im}(P) \cap \mathrm{Im}(I - P)$. Then there exists $y, z \in \mathbb{R}^U$ such that $x = Py$ and $x = (I - P)z$. Therefore

$$x = Py = P^2 y = P(I - P)z = 0.$$

Therefore, $\mathrm{Im}(P) \cap \mathrm{Im}(I - P) = \{0\}$.

To prove the converse, assume then that $\mathrm{Im}(P) \cap \mathrm{Im}(I - P) = \{0\}$. Let $x \in \mathbb{R}^U$. Then

$$P(I - P)x = (I - P)Px.$$

Since the LHS is in $\mathrm{Im}(P)$ and the RHS is in $\mathrm{Im}(I - P)$, it follows that $P(I - P)x = (I - P)Px = 0$. We conclude that $P^2 = P$ if and only if $\mathbb{R}^U = \mathrm{Im}(P) \oplus \mathrm{Im}(I - P)$.

To finish the proof, suffices to show that whenever P is a projector, $\mathrm{Null}(P) = \mathrm{Im}(I - P)$. Since $P^2 = P$, it follows that $P(I - P) = 0$, so that $\mathrm{Im}(I - P) \subseteq \mathrm{Null}(P)$.

If $x \in \mathrm{Null}(P)$, then $x = (I - P)x$, so that $x \in \mathrm{Im}(I - P)$. Therefore $\mathrm{Null}(P) = \mathrm{Im}(I - P)$, and the proof is finished. □

Given a linear subspace S, there are many T such that $\mathbb{R}^U = S \oplus T$. Therefore, in general, there are several projections on a single space S. This issue can be solved exploring the Euclidean structure of the vector space.

Definition 1.37. Let $S \subseteq \mathbb{R}^U$ be a subspace. The *orthogonal projection* on S, denoted P_S, is the projection on S along $S^\perp$.

Proposition 1.38. Let $S \subseteq \mathbb{R}^U$ be a subspace. Then $P \in \mathbb{R}^{U \times U}$ is the orthogonal projector on S if and only if

(i) $\mathrm{Im}(P) = S$,

(ii) $P^2 = P$, and

(iii) $P^\mathsf{T} = P$.

Proof. If $P^2 = P$, Proposition 1.36 ensures that P is the projector on $\mathrm{Im}(P)$ along $\mathrm{Null}(P)$. Since $P^\mathsf{T} = P$, we have that $\mathrm{Null}(P) = \mathrm{Null}(P^\mathsf{T}) = \mathrm{Im}(P)^\perp$, so that (i), (ii) and (iii) hold.

Suppose P is the orthogonal projector on S. Proposition 1.36 ensures that (i) and (ii) holds, and also that $\mathrm{Im}(P) = S$ and $\mathrm{Null}(P) = S^\perp$. Therefore,

$$\mathrm{Im}(P^\mathsf{T}) = \mathrm{Null}(P)^\perp = S,$$
$$\mathrm{Null}(P^\mathsf{T}) = \mathrm{Im}(P)^\perp = S^\perp.$$

Moreover P^T is a projector, since $(P^\mathsf{T})^2 = (P^2)^\mathsf{T} = P^\mathsf{T}$. Therefore P^T is a projector on S along $S^\perp$, so that (iii) holds. □

1.5 The Moore-Penrose Pseudoinverse

Proposition 1.39. Let U and V be finite sets. Assume $A \in \mathbb{R}^{V\times U}$ and $B \in \mathbb{R}^{U\times V}$ are matrices such that $ABA = A$. Then

(i) $\operatorname{Null}(A) = \operatorname{Null}(BA)$, and

(ii) $\operatorname{Im}(A) = \operatorname{Im}(AB)$.

Proof. Since $A = ABA$, item (i) in Proposition 1.29 implies

$$\operatorname{Null}(A) \subseteq \operatorname{Null}(BA) \subseteq \operatorname{Null}(ABA) = \operatorname{Null}(A).$$

In other words, $\operatorname{Null}(AB) = \operatorname{Null}(A)$. Similarly item (ii) in Propositon 1.29 implies

$$\operatorname{Im}(A) = \operatorname{Im}(ABA) \subseteq \operatorname{Im}(AB) \subseteq \operatorname{Im}(A).$$

Hence, $\operatorname{Im}(A) = \operatorname{Im}(AB)$. □

Lemma 1.40. Let U and V be finite sets. Let $A\mathbb{R}^{V\times U}$. Then

(i) $A = P_{\operatorname{Im}(A)}A$,

(ii) $A = AP_{\operatorname{Null}(A)^\perp}$.

Proof. We first prove (ii). Note that $\mathbb{R}^U = \operatorname{Null}(A) \oplus \operatorname{Null}(A)^\perp$. Hence, $I = P_{\operatorname{Null}(A)} + P_{\operatorname{Null}(A)^\perp}$. Therefore,

$$A = A\big(P_{\operatorname{Null}(A)} + P_{\operatorname{Null}(A)^\perp}\big) = AP_{\operatorname{Null}(A)} + AP_{\operatorname{Null}(A)^\perp}.$$

Proposition 1.31 ensures that $AP_{\operatorname{Null}(A)} = 0$, which implies (ii)

Note that (ii) applied to A^T ensures that $A^\mathsf{T} = A^\mathsf{T}P_{\operatorname{Null}(A^\mathsf{T})^\perp}$. Theorem 1.30 ensures that $P_{\operatorname{Null}(A^\mathsf{T})^\perp} = P_{\operatorname{Im}(A)}$. Hence

$$A = (A^\mathsf{T})^\mathsf{T} = (A^\mathsf{T}P_{\operatorname{Im}(A)})^\mathsf{T} = P_{\operatorname{Im}(A)}A.$$ □

Definition 1.41 (Moore-Penrose pseudoinverse)**.** Let U and V be finite sets. Let $A \in \mathbb{R}^{V\times U}$. A *(Moore-Penrose) pseudoinverse of* A is a matrix $A^\dagger \in \mathbb{R}^{U\times V}$ such that

(i) $AA^\dagger = P_{\operatorname{Im}(A)}$, and

(ii) $A^\dagger A = P_{\operatorname{Im}(A^\dagger)}$.

Let $A \in \mathbb{R}^{V\times U}$, with U and V finite. If there exists a pseudoinverse $A^\dagger \in \mathbb{R}^{U\times V}$, since the definition of the pseudoinverse is symmetric on A and $A^\dagger$, it holds that A is the pseudoinverse of $A^\dagger$. Moreover, if A has a inverse, then A^{-1} is a pseudoinverse of A. Furthermore, properties (i) and (ii) of the definition, together with Lemma 1.40 imply

$$AA^\dagger A = P_{\operatorname{Im}(A)}A = A, \tag{1.42}$$

$$A^\dagger AA^\dagger = P_{\operatorname{Im}(A^\dagger)}A^\dagger = A^\dagger. \tag{1.43}$$

Proposition 1.44. Let U and V be finite sets. Let $A \in \mathbb{R}^{V\times U}$, and let $A^\dagger \in \mathbb{R}^{U\times V}$ be a pseudoinverse of A. Then

(i) $\operatorname{Null}(A^\dagger) = \operatorname{Im}(A)^\perp$, and

(ii) $\operatorname{Im}(A^\dagger) = \operatorname{Null}(A)^\perp$.

Proof. Equation (1.43) and item (i) of Proposition 1.39 ensure that $\operatorname{Null}(A^\dagger) = \operatorname{Null}(AA^\dagger)$. Since $AA^\dagger = P_{\operatorname{Im}(A)}$, we have that $\operatorname{Null}(A^\dagger) = \operatorname{Im}(A)^\perp$. Likewise, Equation (1.42) and the same item (i) of Proposition 1.39 ensure that $\operatorname{Null}(A) = \operatorname{Null}(A^\dagger A)$. Since $A^\dagger A = P_{\operatorname{Im}(A^\dagger)}$, we have that $\operatorname{Null}(A) = \operatorname{Im}(A^\dagger)^\perp$. Theorem 1.30 then implies item (ii). □

Proposition 1.45. Let U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$. If A has a pseudoinverse, then it is unique.

Proof. Suppose B and C are two pseudoinverses of A. Proposition 1.44 implies that $\operatorname{Im}(B) = \operatorname{Im}(C)$. Hence, Equation (1.43) and Lemma 1.40 imply

$$B = BAB = BP_{\operatorname{Im}(A)} = BAC = P_{\operatorname{Im}(B)}C = P_{\operatorname{Im}(C)}C = C. \qquad \square$$

Let U and V be finite sets, and let $A \in \mathbb{R}^{V \times U}$. A linear transformation is injective if and only if its nullspace is $\{0\}$. Therefore, for the linear transformation $A\cdot \colon \mathbb{R}^U \to \mathbb{R}^V$, its restriction to $\operatorname{Null}(A)^\perp$ is injective. Hence, it has an inverse

$$\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1} \colon \operatorname{Im}(A) \to \operatorname{Null}(A)^\perp.$$

Since $\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1}$ is the inverse of a linear transformation, it is a linear transformation. Denote by $I_{\operatorname{Im}(A)}$ the identity in the vector space $\operatorname{Im}(A) \subseteq \mathbb{R}^V$, and by $I_{\operatorname{Null}(A)^\perp}$ the identity in $\operatorname{Null}(A)^\perp \subseteq \mathbb{R}^U$. Then

$$\begin{aligned} \left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1} &= I_{\operatorname{Im}(A)}, \\ \left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1}\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right) &= I_{\operatorname{Null}(A)^\perp}. \end{aligned} \tag{1.46}$$

Note that we are no longer dealing with matrices, but with linear operators. However, this procedure works in general, and with such operators it is possible to define a pseudoinverse matrix for every matrix.

Proposition 1.47. Let U and V be finite sets, and let $A \in \mathbb{R}^{V \times U}$. Then A has a pseudoinverse $A^\dagger \in \mathbb{R}^{U \times V}$, and

$$A^\dagger \cdot = \left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1} P_{\operatorname{Im}(A)} \cdot .$$

Proof. Note that the RHS of the equation is defined for every matrix A. Moreover, for every linear transformation from $\mathbb{R}^U$ into $\mathbb{R}^V$, there is a unique matrix such that its operator is said linear transformation. Hence, it is enough to show that the operator on the RHS of the statement is the operator of a pseudoinverse of A.

Denote by B the linear operator $\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)^{-1}$. Since B is surjective in $\operatorname{Null}(A)^\perp$ and $P_{\operatorname{Im}(A)}\cdot$ is surjective in $\operatorname{Im}(A)$, it holds that $\operatorname{Im}(BP_{\operatorname{Im}(A)}) = \operatorname{Im}(B) = \operatorname{Null}(A)^\perp$.

Hence, Lemma 1.40 and Equation (1.46) imply

$$B\big(P_{\operatorname{Im}(A)}A\big)\cdot = B\big(AP_{\operatorname{Null}(A)^\perp}\big)\cdot = B\left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)P_{\operatorname{Null}(A)^\perp}\cdot = I_{\operatorname{Null}(A)^\perp}P_{\operatorname{Null}(A)^\perp}\cdot = P_{\operatorname{Null}(A)^\perp}\cdot .$$

Once again, Lemma 1.40 and Equation (1.46) imply

$$A \cdot B\big(P_{\operatorname{Im}(A)}\big)\cdot = \left(A \cdot \restriction_{\operatorname{Null}(A)^\perp}\right)BP_{\operatorname{Im}(A)}\cdot = I_{\operatorname{Im}(A)}P_{\operatorname{Im}(A)}\cdot = P_{\operatorname{Im}(A)}. \qquad \square$$

Proposition 1.48. Let U and V be finite sets, and let $A \in \mathbb{R}^{V \times U}$. Then

$$(A^\mathsf{T})^\dagger = (A^\dagger)^\mathsf{T}.$$

Proof. Proposition 1.44 ensures $\operatorname{Im}(A^\dagger) = \operatorname{Im}(A^\mathsf{T})$. Also, by definition, $A^\dagger A = P_{\operatorname{Im}(A^\dagger)}$. Hence

$$A^\mathsf{T}(A^\dagger)^\mathsf{T} = (A^\dagger A)^\mathsf{T} = P^\mathsf{T}_{\operatorname{Im}(A^\dagger)} = P_{\operatorname{Im}(A^\dagger)} = P_{\operatorname{Im}(A^\mathsf{T})}.$$

Moreover, Proposition 1.44 implies $\operatorname{Im}(A) = \operatorname{Im}((A^\dagger)^\mathsf{T})$. Also, $AA^\dagger = P_{\operatorname{Im}(A)}$. Then

$$(A^\dagger)^\mathsf{T}A^\mathsf{T} = (AA^\dagger)^\mathsf{T} = P^\mathsf{T}_{\operatorname{Im}(A)} = P_{\operatorname{Im}(A)} = P_{\operatorname{Im}((A^\dagger)^\mathsf{T})}.$$

In other words, $(A^\dagger)^\mathsf{T}$ is the pseudoinverse of A^T. $\square$

Corollary 1.49. Let U be a finite set. If $A \in \mathbb{R}^{U\times U}$ is a symmetric matrix, then

$$AA^\dagger = A^\dagger A.$$

Proof. Use that $P_{\mathrm{Im}(A)}$ is symmetric and Proposition 1.48:

$$AA^\dagger = P_{\mathrm{Im}(A)} = P_{\mathrm{Im}(A)}^\mathsf{T} = (AA^\dagger)^\mathsf{T} = (A^\dagger)^\mathsf{T} A^\mathsf{T} = (A^\mathsf{T})^\dagger A^\mathsf{T} = A^\dagger A. \qquad \square$$

Unfortunately, it is not always the case that $(AB)^\dagger = B^\dagger A^\dagger$. One of the notable cases when this holds is the following.

Proposition 1.50. Let U, V and T be finite sets, and let $A \in \mathbb{R}^{V\times U}$ and $B \in \mathbb{R}^{U\times T}$. If A is injective and B is surjective, then

$$(AB)^\dagger = B^\dagger A^\dagger.$$

Proof. Since A is injective, Proposition 1.44 and Theorem 1.30 imply

$$\mathrm{Im}(A^\dagger) = \mathrm{Im}(A^\mathsf{T}) = \mathrm{Null}(A)^\perp = \{0\}^\perp = \mathbb{R}^U.$$

Therefore, $A^\dagger A = I_U$. Hence

$$B^\dagger A^\dagger AB = B^\dagger I_U B = B^\dagger B = P_{\mathrm{Im}(B^\dagger)}.$$

Moreover, note that Theorem 1.30 implies that A^T is surjective. Furthermore, Proposition 1.44 ensures

$$\mathrm{Im}((AB)^\dagger) = \mathrm{Im}((AB)^\mathsf{T}) = \mathrm{Im}(B^\mathsf{T} A^\mathsf{T}) = \mathrm{Im}(B^\mathsf{T}) = \mathrm{Im}(B^\dagger).$$

Theorefore, $B^\dagger A^\dagger AB = P_{\mathrm{Im}((AB)^\dagger)}$. It remains to prove that $ABB^\dagger A^\dagger = P_{\mathrm{Im}(AB)}$. Since B is surjective, we have that $\mathrm{Im}(AB) = \mathrm{Im}(A)$, and that $P_{\mathrm{Im}(B)} = I_U$. Hence

$$ABB^\dagger A^\dagger = AI_U A^\dagger = AA^\dagger = P_{\mathrm{Im}(A)} = P_{\mathrm{Im}(AB)}. \qquad \square$$

Proposition 1.51. Let U be a finite set. Then for $\mathbb{1} \in \mathbb{R}^U$ and $P_{\mathrm{span}(\mathbb{1})} \in \mathbb{R}^{U\times U}$,

$$P_{\mathrm{span}(\mathbb{1})} = \frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}.$$

Proof. Suffices to show that the RHS of the statement is the orthogonal projector on $\mathrm{span}(\mathbb{1})$. First, note that

$$\left(\frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}\right)^\mathsf{T} = \frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T},$$

so that $(1/|U|)\mathbb{1}\mathbb{1}^\mathsf{T}$ is symmetric. Moreover,

$$\left(\frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}\right)^2 = \frac{\mathbb{1}^\mathsf{T}\mathbb{1}}{|U|^2}\mathbb{1}\mathbb{1}^\mathsf{T} = \frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}.$$

Hence, $(1/|U|)\mathbb{1}\mathbb{1}^\mathsf{T}$ is a projector. To conclude the proof, it remains to show that $\mathrm{span}(\mathbb{1}) = \mathrm{Im}(1/|U|\,\mathbb{1}\mathbb{1}^\mathsf{T})$. Let $x \in \mathrm{span}(\mathbb{1})$. Then there exists $\alpha \in \mathbb{R}$ such that $x = \alpha\mathbb{1}$. Hence

$$\left(\frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}\right)x = \frac{\mathbb{1}^\mathsf{T}x}{|U|}\mathbb{1} = \frac{\alpha\mathbb{1}^\mathsf{T}\mathbb{1}}{|U|}\mathbb{1} = \alpha\frac{|U|}{|U|}\mathbb{1} = \alpha\mathbb{1} = x.$$

In other words, $\mathrm{span}(\mathbb{1}) \subseteq \mathrm{Im}((1/|U|)\mathbb{1}\mathbb{1}^\mathsf{T})$. Let $y \in \mathbb{R}^U$. Then

$$\left(\frac{1}{|U|}\mathbb{1}\mathbb{1}^\mathsf{T}\right)y = \left(\frac{\mathbb{1}^\mathsf{T}y}{|U|}\right)\mathbb{1}.$$

In other words, $\mathrm{Im}((1/|U|)\mathbb{1}\mathbb{1}^\mathsf{T}) \subseteq \mathrm{span}(\mathbb{1})$, and the proof is done. $\square$

Example 1.52. Let V be a finite set, and let $i \in V$. Set $U \coloneqq V \setminus \{i\}$, and let $A \in \mathbb{R}^{V \times U}$ be defined as

$$\begin{aligned} A[U,U] &= I, \\ A[i,U] &= -\mathbb{1}^{\mathsf{T}}. \end{aligned}$$

In other words, assuming the topmost rows are indexed by U,

$$A = \begin{bmatrix} I \\ -\mathbb{1}^{\mathsf{T}} \end{bmatrix}.$$

We wish to calculate $A^{\dagger}$. For $y \in \mathbb{R}^V$, we have that $y \perp \mathbb{1}$ if and only if there exists $x \in \mathbb{R}^U$ such that

$$y = \begin{bmatrix} x \\ -\mathbb{1}^{\mathsf{T}} x \end{bmatrix} = \begin{bmatrix} I \\ -\mathbb{1}^{\mathsf{T}} \end{bmatrix} x.$$

Hence, $\operatorname{Im}(A) = \operatorname{span}(\mathbb{1})^{\perp}$. Moreover, since $I = P_{\operatorname{span}(\mathbb{1})} + P_{\operatorname{span}(\mathbb{1})^{\perp}}$, Proposition 1.51 ensures that

$$P_{\operatorname{span}(\mathbb{1})^{\perp}} = I - \frac{1}{|V|} \mathbb{1}\mathbb{1}^{\mathsf{T}}.$$

Therefore, since $AA^{\dagger} = P_{\operatorname{Im}(A)}$, we conclude that

$$\begin{bmatrix} I \\ -\mathbb{1}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} A^{\dagger}[U,U] & A^{\dagger}[U,i] \end{bmatrix} = \begin{bmatrix} A^{\dagger}[U,U] & A^{\dagger}[U,i] \\ \mathbb{1}^{\mathsf{T}} A^{\dagger}[U,U] & \mathbb{1}^{\mathsf{T}} A^{\dagger}[U,i] \end{bmatrix} = \begin{bmatrix} I - \frac{1}{|V|}\mathbb{1}\mathbb{1}^{\mathsf{T}} & -\frac{1}{|V|}\mathbb{1} \\ -\frac{1}{|V|}\mathbb{1}^{\mathsf{T}} & 1 - \frac{1}{|V|} \end{bmatrix}.$$

Therefore, assuming then that the leftmost columns of $A^{\dagger}$ are indexed by U, we conclude that

$$A^{\dagger} = \begin{bmatrix} I - \frac{1}{|V|}\mathbb{1}\mathbb{1}^{\mathsf{T}} & -\frac{1}{|V|}\mathbb{1} \end{bmatrix}.$$

1.6 The Laplacian

Definition 1.53. Let $D = (V, A, \psi)$ be a digraph. The *head matrix of* D is the matrix defined as

$$H_D \coloneqq \sum_{ij \in A} e_j e_{ij}^{\mathsf{T}}.$$

Moreover, the *tail matrix of* D is the matrix defined as

$$T_D \coloneqq \sum_{ij \in A} e_i e_{ij}^{\mathsf{T}}.$$

Let $D = (V, A, \psi, w)$ be a digraph. Both matrices H_D and T_D enable us to define several others interesting matrices related to D. The *incidence matrix of* D is the matrix defined as

$$B_D \coloneqq H_D - T_D.$$

The *adjencency matrix of* D is the matrix defined as

$$A_D \coloneqq H_D \operatorname{Diag}(w) T_D^{\mathsf{T}}.$$

The *degree matrix of* D is the matrix defined as

$$D_D \coloneqq H_D \operatorname{Diag}(w) H_D^{\mathsf{T}}.$$

Finally, the *Laplacian of* D is the matrix defined as

$$L_D \coloneqq H_D \operatorname{Diag}(w) B_D^{\mathsf{T}}.$$

Note that it is quite simple to relate the matrices involving D^{T} with the ones related to D by the following equalities:

$$\begin{aligned} H_{D^{\mathsf{T}}} &= T_D, \\ T_{D^{\mathsf{T}}} &= H_D. \end{aligned}$$

Hence, we have both that $B_{D^{\mathsf{T}}} = -B_D$ and the pleasant equality $A_{D^{\mathsf{T}}} = A_D^{\mathsf{T}}$.

Proposition 1.54. Let $D = (V, A, \psi)$ be a digraph. Then

$$H_D^{\mathsf{T}} \mathbb{1} = T_D^{\mathsf{T}} \mathbb{1} = \mathbb{1}.$$

Proof. For every $k \in V$, we have that $1 = \sum_{i \in V} [k = i]$. Therefore, using the definitions o H_D and $\mathbb{1}$,

$$H_D^{\mathsf{T}} \mathbb{1} = \left(\sum_{jk \in A} e_k e_{jk}^{\mathsf{T}} \right)^{\mathsf{T}} \left(\sum_{i \in V} e_i \right) = \sum_{jk \in A} \sum_{i \in V} e_{jk} e_k^{\mathsf{T}} e_i = \sum_{jk \in A} \sum_{i \in V} [k = i] e_{jk} = \sum_{jk \in A} e_{jk} = \mathbb{1}.$$

Since the above arguments holds for any digraph, and $T_D = H_{D^{\mathsf{T}}}$, the proof is finished. □

Proposition 1.55. Let $D = (V, A, \psi, w)$ be a weighted digraph. Then $D_D \mathbb{1} = A_D \mathbb{1}$, and for every $i \in V$,

$$e_i^{\mathsf{T}} A_D \mathbb{1} = \sum w(\delta^{\text{in}}(i)).$$

Proof. Proposition 1.54 implies

$$D_D \mathbb{1} = H_D \operatorname{Diag}(w) H_D^{\mathsf{T}} \mathbb{1} = H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} \mathbb{1} = A_D \mathbb{1}.$$

Moreover, for every $i \in V$, Proposition 1.54 ensures

$$\begin{aligned} e_i^{\mathsf{T}} A_D \mathbb{1} &= e_i^{\mathsf{T}} H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} \mathbb{1} = e_i^{\mathsf{T}} H_D \operatorname{Diag}(w) \mathbb{1} = e_i^{\mathsf{T}} H_D \left(\sum_{jk \in A} w(jk) e_{jk} \right) \\ &= \sum_{jk \in A} w(jk) e_i^{\mathsf{T}} H_D e_{jk} = \sum_{jk \in A} w(jk) e_i^{\mathsf{T}} e_k = \sum_{jk \in A} [k = i] w(jk) \\ &= \sum w(\delta^{\text{in}}(i)). \end{aligned}$$ □

Let G be a graph. We define the adjacency, degree, incidence and Laplacian of G as the corresponding matrix for the symmetric digraph of G.

Proposition 1.56. Let $G = (V, E, \psi, w)$ be a weighted graph. Then

(i) If D is the symmetric digraph of G, then $A_G = A_D$ and $D_G = D_D$.

(ii) For any orientation D of G, it holds that $A_G = H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) H_D^{\mathsf{T}}$.

(iii) $A_G^{\mathsf{T}} = A_G$.

(iv) For any orientation D of G, it holds that $D_G = H_D \operatorname{Diag}(w) H_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) T_D^{\mathsf{T}}$.

(v) For every $i, j \in V$, we have that $(A_G)_{ij} = [ij \in E] w(ij)$.

Proof. Item (i) is the definition of A_G and D_G restated.

Suppose then that $D = (V, A, \phi, w)$ is an orientation of D. If we index the leftmost columns of E by A, we have that

$$\begin{aligned} H_G &= \begin{bmatrix} H_D & H_{D^{\mathsf{T}}} \end{bmatrix} = \begin{bmatrix} H_D & T_D \end{bmatrix}, \\ T_G &= \begin{bmatrix} T_D & T_{D^{\mathsf{T}}} \end{bmatrix} = \begin{bmatrix} T_D & H_D \end{bmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} A_G &= H_G \operatorname{Diag}(w) T_G^{\mathsf{T}} \\ &= \begin{bmatrix} H_D & T_D \end{bmatrix} \begin{bmatrix} \operatorname{Diag}(w) & 0 \\ 0 & \operatorname{Diag}(w) \end{bmatrix} \begin{bmatrix} T_D^{\mathsf{T}} \\ H_D^{\mathsf{T}} \end{bmatrix} \\ &= H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) H_D^{\mathsf{T}}. \end{aligned}$$

Hence, item (ii) holds, with item (iii) is a consequence. Item (iv) follows from a similiar argument:

$$\begin{aligned} D_G &= H_G \operatorname{Diag}(w) H_G^{\mathsf{T}} \\ &= \begin{bmatrix} H_D & T_D \end{bmatrix} \begin{bmatrix} \operatorname{Diag}(w) & 0 \\ 0 & \operatorname{Diag}(w) \end{bmatrix} \begin{bmatrix} H_D^{\mathsf{T}} \\ T_D^{\mathsf{T}} \end{bmatrix} \\ &= H_D \operatorname{Diag}(w) H_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) T_D^{\mathsf{T}}. \end{aligned}$$

Finally, to prove (v), let $D = (V, A, \phi, w)$ be any orientation of G. Note that

$$\begin{aligned} H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} &= H_D \operatorname{Diag}(w) \left(\sum_{ij \in A} e_{ij} e_i^{\mathsf{T}} \right) \\ &= \sum_{ij \in A} H_D \operatorname{Diag}(w) e_{ij} e_i^{\mathsf{T}} \\ &= \sum_{ij \in A} w(ij) H_D e_{ij} e_i^{\mathsf{T}} \\ &= \sum_{ij \in A} w(ij) e_j e_i^{\mathsf{T}}. \end{aligned}$$

Therefore, $(H_D \operatorname{Diag}(w) T_D^{\mathsf{T}})_{ij} = [ij \in A] w(ij)$. Similarly, we have that $(T_D \operatorname{Diag}(w) H_D^{\mathsf{T}})_{ij} = [ji \in A] w(ij)$. Hence, item (ii) and the fact that D is an orientation of G finish the proof

$$(A_G)_{ij} = [ij \in A] w(ij) + [ji \in A] w(ij) = [ij \in E] w(ij). \qquad \square.$$

Proposition 1.57. Let $G = (V, E, \psi, w)$ be a graph. Let L_G be the Laplacian of G. Then

(i) If D is the symmetric graph of G, then $L_G = L_D$.

(ii) $L_G = D_G - A_G$,

(iii) For any orientation D of G, we have that $L_G = B_D \operatorname{Diag}(w) B_D^{\mathsf{T}}$.

(iv) $L_G^{\mathsf{T}} = L_G$.

(v) $L_G = \sum_{ij \in E} w(ij)(e_j - e_i)(e_j - e_i)^{\mathsf{T}}$,

Proof. Item (i) is the definition of L_G.

Item (ii) follows from the following calculations:

$$L_G = H_G \operatorname{Diag}(w) B_G^{\mathsf{T}} = H_G \operatorname{Diag}(w)(H_G - T_G)^{\mathsf{T}} = H_G \operatorname{Diag}(w) H_G^{\mathsf{T}} - H_G \operatorname{Diag}(w) T_G^{\mathsf{T}} = D_G - A_G.$$

Let D be an orientation of G. Proposition 1.56 ensures

$$\begin{aligned} L_G &= D_G - A_G \\ &= H_D \operatorname{Diag}(w) H_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) T_D^{\mathsf{T}} - \big(H_D \operatorname{Diag}(w) T_D^{\mathsf{T}} + T_D \operatorname{Diag}(w) H_D^{\mathsf{T}} \big) \\ &= H_D \operatorname{Diag}(w) \big(H_D^{\mathsf{T}} - T_D^{\mathsf{T}} \big) + T_D \operatorname{Diag}(w) \big(T_D^{\mathsf{T}} - H_D^{\mathsf{T}} \big) \\ &= H_D \operatorname{Diag}(w) B_D^{\mathsf{T}} - T_D \operatorname{Diag}(w) B_D^{\mathsf{T}} \\ &= B_D \operatorname{Diag}(w) B_D^{\mathsf{T}}. \end{aligned}$$

Hence, item (iii) holds, and item (iv) is a consequence.

For item (v), let D be any orientation of G. Then item (iii) ensures that $L_G = B_D \operatorname{Diag}(w) B_D^\mathsf{T}$. Hence

$$\begin{aligned}
L_G &= B_D \operatorname{Diag}(w) \left(\sum_{ij \in E} (e_j - e_i) e_{ij}^\mathsf{T} \right)^\mathsf{T} \\
&= B_D \operatorname{Diag}(w) \left(\sum_{ij \in E} e_{ij} (e_j - e_i)^\mathsf{T} \right) \\
&= \sum_{ij \in E} B_D \operatorname{Diag}(w) e_{ij} (e_j - e_i)^\mathsf{T} \\
&= \sum_{ij \in E} w(ij) B_D e_{ij} (e_j - e_i)^\mathsf{T} \\
&= \sum_{ij \in E} w(ij) (e_j - e_i)(e_j - e_i)^\mathsf{T}. \qquad \square
\end{aligned}$$

Proposition 1.58. Let $G = (V, E, w)$ be a connected graph, with $w \in \mathbb{R}^V_{++}$. Then

$$L_G L_G^\dagger = L_G^\dagger L_G = P_{\operatorname{span}(\mathbb{1})^\perp}.$$

Proof. Let D be an orientation of G. Proposition 1.57 ensures $L_G = B_D W B_D^\mathsf{T}$. Hence, L_G is symmetric. Corollary 1.49, Proposition 1.44 and Theorem 1.30 imply that

$$L_G L_G^\dagger = L_G^\dagger L_G = P_{\operatorname{Im}(L_G^\dagger)} = P_{\operatorname{Im}(L_G^\mathsf{T})} = P_{\operatorname{Null}(L_G)^\perp}.$$

It suffices to show that $\operatorname{Null}(L_G) = \operatorname{span}(\mathbb{1})$. Since G is connected, Proposition 2.25 ensures $\operatorname{Null}(B_D^\mathsf{T}) = \operatorname{span}(\mathbb{1})$. It is then enough to prove that $\operatorname{Null}(L_G) = \operatorname{Null}(B_D^\mathsf{T})$.

Note that since $L_G = B_D W B_D^\mathsf{T}$, Proposition 1.29 ensures that $\operatorname{Null}(B_D^\mathsf{T}) \subseteq \operatorname{Null}(L_G)$. Let then $x \in \operatorname{Null}(L_G)$. Then $x^\mathsf{T} L_G x = 0$. Therefore,

$$0 = x^\mathsf{T} L_G x = x^\mathsf{T} B_D W B_D^\mathsf{T} x = \big(B_D^\mathsf{T} x\big) W \big(B_D^\mathsf{T} x\big) = \sum_{ij \in E} w(ij) \big(B_D^\mathsf{T} x\big)_{ij}^2.$$

Since $w(ij) > 0$ for every $ij \in E$, this implies that $(B_D^\mathsf{T} x)_{ij} = 0$ for every $ij \in E$. Hence, $x \in \operatorname{Null}(B_D^\mathsf{T})$. $\square$

1.7 Harmonic Functions

Definition 1.59. Let $D = (V, A, w)$ be a weighted digraph, with $w \in \mathbb{R}^A_{++}$. A function $f \in \mathbb{R}^V$ is *harmonic*, with respect to D, at $i \in V$ if

$$f(i) = \sum_{ik \in A} \frac{w(ik)}{w(i)} f(k).$$

The function f is said to be harmonic in a set $S \subseteq V$ if it is harmonic in every vertex in S.

Proposition 1.60. Let $D = (V, A, w)$ be a weighted digraph, with $w \in \mathbb{R}^A_{++}$. Let $f \in \mathbb{R}^V$ be harmonic in $S \subseteq V$. Suppose further that f reaches its maximum in a vertex $i \in S$. Then, if there is an ij-walk in D such that either every vertex is in S, or every vertex but j is in S, we have that

$$f(i) = f(j).$$

Proof. The proof is by induction on the lenght of the walk. Let $(u_0, \ldots, u_m)$ be an ij-walk, with $u_0 = i$, $u_m = j$, and $u_i \in S$ for every $i \in [m]$.

If $m = 0$, then $i = j$ and the thesis holds.

If $m > 0$, denote by k the vertex u_1. Suppose $f(k) < f(i)$. Since $ik \in A$ and $f(i)$ is a maximum of f, we have that

$$\sum_{ij\in A} \frac{w(ij)}{w(i)} f(j) < \sum_{ij\in A} \frac{w(ij)}{w(i)} f(i) = \left(\sum_{i\in A} \frac{w(ij)}{w(i)} \right) f(i) = 1 \cdot f(i) = f(i).$$

This contradicts the fact that f is harmonic at i. Hence, $f(i) \leq f(k)$. Since $f(i)$ is maximum, we have that $f(i) = f(k)$. This, in turns ensures that the induction hypothesis apply to the kj-walk, so that

$$f(i) = f(k) = f(j). \qquad \square$$

Definition 1.61. Let $G = (V, E, w)$ be a weighted graph, with $w \in \mathbb{R}^A_{++}$. A function $f \in \mathbb{R}^V$ is *harmonic* at $i \in V$ if it is hamornic in the symmetric digraph of G. The function f is said to be harmonic in a set $S \subseteq V$ if it is harmonic in every vertex in S.

Proposition 1.62. Let $D = (V, A, w)$ be a weighted digraph, with $w \in \mathbb{R}^A_{++}$. Suppose D is strongly connected. A function $f \colon V \to \mathbb{R}$ that is harmonic in at least $|V| - 1$ vertices is constant.

Proof. If $|V| \leq 1$, there is nothing to prove. Let $i \in V$ be the vertex in which $f(i)$ is maximum. Denote by $r \in V$ the vertex in which f is not harmonic. For every vertex $j \in V$ there is an ij-walk such that every internal vertex is in $V \setminus \{r\}$. Proposition 1.60 then ensures that $f(i) = f(j)$. □

Proposition 1.63. Let $G = (V, E, w)$ be a weighted graph, with $w \in \mathbb{R}^A_{++}$. Suppose G is connected. Then a function f that is harmonic in every vertex of G is constant.

Proof. Since G is connected, its symmetric digraph is strongly connected. Proposition 1.62 finishes the proof. □

1.8 Extended Positive Reals

The main point of the current section is to justify the manipulations on series of numbers in $\mathbb{R}_+ \cup \{\infty\}$ that will be used in the work. It is not an attempt to develop a whole theory of convergence on the real line. The work in this section is based on the first remarks made by Tao in [19].

The *real numbers*, denote by $\mathbb{R}$, are assumed to have a total order $\leq$, such that for all real numbers $\alpha, \beta, \gamma \in \mathbb{R}$ it holds that

1. if $\alpha \leq \beta$, then $\alpha + \gamma \leq \beta + \gamma$, and
2. if $0 \leq \alpha$ and $0 \leq \beta$, then $0 \leq \alpha\beta$.

Given a subset $S \subseteq \mathbb{R}$, a real number $\alpha \in \mathbb{R}$ is said to be an *upper bound of* S if for every $s \in S$ it holds that $s \leq \alpha$. Similarly, a real number $\alpha \in \mathbb{R}$ is said to be a *lower bound of* S if for every $s \in S$ it holds that $\alpha \leq s$. The real numbers are assumed to be *complete*, meaning that for every nonempty subset $S \subseteq \mathbb{R}$, if S has a lower bound, then S has a greatest lower bound, denoted by $\inf S$. Similarly, if S is nonempty and has an upper bound, it has a least upper bound, denoted by $\sup S$. For a given subset $S \subseteq \mathbb{R}$, the numbers $\inf S$ and $\sup S$ are said to be the *infimum of* S and *supremum of* S, respectively.

Let $S \subseteq \mathbb{R}$ be nonempty. Suppose S has an upper bound. Then $\alpha = \sup S$ if and only if for every $\varepsilon > 0$ there exists $s \in S$ such that

$$\alpha - \varepsilon < s \leq \alpha.$$

The fact that $s \leq \alpha$ is a consequence of α being an upper bound of S, and the existence of $s \in S$ such that $\alpha - \varepsilon < s$ is a consequence of α being the least upper bound. If for every $s \in S$ this inequality did not hold, then $\alpha - \varepsilon$ would be an upper bound for S, which is smaller than α.

A *real sequence* is a function $\alpha \colon \mathbb{N} \to \mathbb{R}$. It is usually denoted as $(\alpha_n)_{n\in\mathbb{N}}$. Moreover, for a given $n \in \mathbb{N}$, the value $\alpha(n)$ is usually denoted as α_n. For a given $\alpha \in \mathbb{R}$, a real sequence is said to *converge to* α if for every real $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $|\alpha_n - \alpha| < \varepsilon$. This is denoted as

$$\lim_{n\to\infty} \alpha_n = \alpha.$$

If $(\alpha_n)_{n\in\mathbb{N}}$ converges to $\alpha \in \mathbb{R}$, then α is said to be the *limit of* $(\alpha_n)_{n\in\mathbb{N}}$. A sequence that converges to some value $\alpha \in \mathbb{R}$ is said to be *convergent.* A real sequence $(\alpha_n)_{n\in\mathbb{N}}$ is said to be *increasing* if $i \leq j$ implies that $\alpha_i \leq \alpha_j$. Moreover, the sequence is said to be *bounded* if there exists $\beta \in \mathbb{R}$ such that for every $n \in \mathbb{N}$ we have that $|\alpha_n| \leq \beta$.

Theorem 1.64. Let $(\alpha_n)_{n\in\mathbb{N}}$ be a bounded and increasing real sequence. Then $(\alpha_n)_{n\in\mathbb{N}}$ is convergent, and

$$\lim_{n\to\infty} \alpha_n = \sup\{\, \alpha_n : n \in \mathbb{N}\}.$$

Proof. Set $\alpha \coloneqq \sup\{\, \alpha_n : n \in \mathbb{N}\}$, and let $\varepsilon > 0$. Since α is the supremum of $\{\, \alpha_n : n \in \mathbb{N}\}$, it holds that there exists $N \in \mathbb{N}$ such that

$$\alpha - \varepsilon < \alpha_N.$$

In other words, $\alpha - \alpha_N < \varepsilon$. Moreover, since $(\alpha_n)_{n\in\mathbb{N}}$ is increasing, for every $n \in \mathbb{N}$ with $N \leq n$ we have that $\alpha_N \leq \alpha_n$. Hence $\alpha - \alpha_n \leq \alpha - \alpha_N$. However, we also have that $0 \leq \alpha - \alpha_n$, since α was defined as the supremum of the values of the sequence. Therefore, for every $n \in \mathbb{N}$ such that $N \leq n$,

$$|\alpha - \alpha_n| = \alpha - \alpha_n \leq \alpha - \alpha_N < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is done. □

The *extended positive reals* is the set $\mathbb{R}_+ \cup \{\infty\}$. We extend the order, the addition and the multiplication on $\mathbb{R}_+$ to $\mathbb{R}_+ \cup \{\infty\}$ by defining addition, multiplication and order for ∞. For every $\alpha \in \mathbb{R}_+ \cup \{\infty\}$, define

$$\begin{aligned} \alpha &\leq \infty, \\ \alpha + \infty &\coloneqq \infty, \\ \infty + \alpha &\coloneqq \infty. \end{aligned}$$

Moreover, for every nonzero $\beta \in \mathbb{R}_+$ define

$$\begin{aligned} \beta\infty &\coloneqq \infty, \\ \infty\beta &\coloneqq \infty. \end{aligned}$$

Also, $0\infty \coloneqq 0$, and $\infty 0 \coloneqq 0$. Note that many operations are not defined in $\mathbb{R}_+ \cup \{\infty\}$ as a whole; mainly substraction and division by ∞. However, the elements of $\mathbb{R}_+ \cup \{\infty\}$ that are not ∞ — called the *finite* elements of $\mathbb{R}_+ \cup \{\infty\}$ – are in $\mathbb{R}$, so that every result about real numbers applies to them.

Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+ \cup \{\infty\}$, i.e., $\alpha \colon \mathbb{N} \to \mathbb{R}_+ \cup \{\infty\}$. If there exists $N \in \mathbb{N}$ such that $N \leq n$ implies that $\alpha_n \in \mathbb{R}_+$, and such that the real sequence $(\beta_n)_{n\in\mathbb{N}}$, defined as $\beta_n \coloneqq \alpha_{n+N}$ converges to α, we say that $(\alpha_n)_{n\in\mathbb{N}}$ converges to α. This is denoted as

$$\lim_{n\to\infty} \alpha_n = \alpha.$$

Moreover, if for every $\beta \in \mathbb{R}_+$ there exists $N \in \mathbb{N}$ such that $N \leq n$ implies that $\beta \leq \alpha_n$, we say that α_n *converges to infinity.* This is denoted as

$$\lim_{n\to\infty} \alpha_n = \infty.$$

The most useful property of the extended positive reals is the fact that every series in it converges. Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+ \cup \{\infty\}$. Define the sequence $(\beta_n)_{n\in\mathbb{N}}$ as

$$\beta_n \coloneqq \sum_{i=0}^{n} \alpha_i,$$

for every $n \in \mathbb{N}$. Since for every n we have that $0 \leq \alpha_n$, it holds that β_n is a increasing sequence. If it is bounded, Theorem 1.64 ensures that β_n converges to a real number. If it is not bounded, then by definition $\lim_{n\to\infty} \beta_n = \infty$. Hence, the limit is always defined, and is always an element in $\mathbb{R}_+ \cup \{\infty\}$. Therefore, for a given sequence of extended real numbers $(\alpha_n)_{n\in\mathbb{N}}$, define

$$\sum_{n\in\mathbb{N}} \alpha_n \coloneqq \lim_{k\to\infty} \sum_{n=0}^{k} \alpha_n.$$

This definition has a couple more convenient equivalences. First, note that when $\sum_{n\in\mathbb{N}} \alpha_n = \alpha$, for some $\alpha \in \mathbb{R}$, this definition is equivalent with the statement that for every $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that $k > K$ implies that

$$\sum_{n=0}^{k} \alpha_n > \alpha - \varepsilon.$$

Moreover, if $\sum_{n\in\mathbb{N}} \alpha_n = \infty$, then for every $L \in \mathbb{R}$ there exists $k \in \mathbb{N}$ such that $k > K$ implies that

$$\sum_{n=0}^{k} \alpha_k > L.$$

Another convenient equivalence of the series summation definition is that, for every $(\alpha_n)_{n\in\mathbb{N}}$,

$$\sum_{n\in\mathbb{N}} \alpha_n = \sup\left\{ \sum_{n=0}^{k} \alpha_n : k \in \mathbb{N} \right\}. \tag{1.65}$$

In particular, this implies the fact that $\alpha_n \leq \sum_{n\in\mathbb{N}} \alpha_n$ for every $n \in \mathbb{N}$.

Proposition 1.66. Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_+ \cup \{\infty\}$. Suppose that $\alpha_n \leq \beta_n$ for every $n \in \mathbb{N}$. Then

$$\sum_{n\in\mathbb{N}} \alpha_n \leq \sum_{n\in\mathbb{N}} \beta_n.$$

Proof. If $\sum_{n\in\mathbb{N}} \beta_n = \infty$, there is nothing to prove. Suppose then that $\sum_{n\in\mathbb{N}} \beta_n = \beta$, for $\beta \in \mathbb{R}$. Then for every $k \in \mathbb{N}$,

$$\sum_{n=0}^{k} \alpha_n \leq \sum_{n=0}^{k} \beta_n \leq \beta.$$

Therefore, $\sum_{n\in\mathbb{N}} \alpha_n$ is bounded, so there exists $\alpha \in \mathbb{R}$ such that $\sum_{n\in\mathbb{N}} \alpha_n = \alpha$. It is enough to show that $\alpha \leq \beta$. Suppose that $\beta < \alpha$. Then there exists $K \in \mathbb{N}$ such that $k > K$ implies that

$$\sum_{n=0}^{k} \alpha_n > \alpha - \frac{\alpha - \beta}{2} = \beta + \frac{\alpha - \beta}{2} \geq \sum_{n=0}^{k} \beta_n.$$

But this implies that there exists $i \in \mathbb{N}$, with $i \leq n$, such that $\alpha_i > \beta_i$. □

Proposition 1.67. Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+ \cup \{\infty\}$. Let $\beta \in \mathbb{R}_+ \cup \{\infty\}$. Then

$$\sum_{n\in\mathbb{N}} \beta\alpha_n = \beta \sum_{n\in\mathbb{N}} \alpha_n.$$

Proof. Set $\alpha \coloneqq \sum_{n\in\mathbb{N}} \alpha_n$. Since the product in $\mathbb{R}_+ \cup \{\infty\}$ was divided into cases, this proof will have to treat each separetely.

If $\alpha = 0$, we have that $\alpha_n = 0$ for every $n \in \mathbb{N}$. Therefore, if either α or β are zero, it holds that $\beta\alpha_n = 0$ for every natural n. Hence

$$\sum_{n\in\mathbb{N}} \beta\alpha_n = 0 = \beta \sum_{n\in\mathbb{N}} \alpha_n.$$

Suppose that $\beta = \infty$ and α is nonzero. In this case, $\beta\alpha_n = \infty$ for every $n \in \mathbb{N}$. Therefore $\sum_{n\in\mathbb{N}} \beta\alpha_n = \infty$.

Suppose than that β is a nonzero real number, and $\alpha = \infty$. Let $L \in \mathbb{R}$ be any. Since β is nonzero, $\beta^{-1} \in \mathbb{R}$. Since $\alpha = \infty$, there exists $K \in \mathbb{N}$ such that $k > K$ implies that $\sum_{n=0}^{k} \alpha_n > L\beta^{-1}$. Therefore, for every $k > K$ we have that

$$\sum_{n=0}^{k} \beta\alpha_n > L.$$

Hence, $\sum_{n\in\mathbb{N}} \beta\alpha_N = \infty$, which is equal to $\beta\alpha$.

Finally, suppose both α and β are nonzero real numbers. Let $\varepsilon > 0$. Since $\sum_{n\in\mathbb{N}} \alpha_n = \alpha$, we have that there exists $K \in \mathbb{N}$ such that $k > K$ implies that $\sum_{n=0}^{k} \alpha_n > \alpha - \varepsilon\beta^{-1}$. Hence, for every $k \geq K$,

$$\sum_{n=0}^{k} \beta\alpha_n > \beta\alpha - \varepsilon. \qquad \square$$

Theorem 1.68. Let V be a finite set. For every $i \in V$, let $((\alpha_i)_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+ \cup \{\infty\}$. Then

$$\sum_{n\in\mathbb{N}}\sum_{i\in V}(\alpha_i)_n = \sum_{i\in V}\sum_{n\in\mathbb{N}}(\alpha_i)_n.$$

Proof. Suppose there exists $j \in V$ such that $\sum_{n\in\mathbb{N}}(\alpha_j)_n = \infty$. Since for every $n \in \mathbb{N}$ we have that $\sum_{i\in V}(\alpha_i)_n \geq (\alpha_j)_n$, Proposition 1.66 ensures

$$\sum_{n\in\mathbb{N}}\sum_{i\in V}(\alpha_i)_n \geq \sum_{n\in\mathbb{N}}(\alpha_j)_n = \infty.$$

Hence, the LHS of the statement is equal to ∞, just like the RHS.

Suppose then that for every $i \in V$, the series $\sum_{n\in\mathbb{N}}(\alpha_i)_n$ is bounded. Then, for every $i \in V$, there exists a real number β_i such that $\sum_{n\in\mathbb{N}}(\alpha_i)_n = \beta_i$. Let $\varepsilon > 0$. For every $i \in V$, there exists $N_i \in \mathbb{N}$ such that $N_i \leq n$ implies that

$$\sum_{k=0}^{n}(\alpha_i)_k > \beta_i - \varepsilon/\left|V\right|.$$

Set $N \coloneqq \max_{i\in V} N_i$. Then for every $n \in \mathbb{N}$ such that $N \leq n$ we have that

$$\sum_{k=0}^{n}\sum_{i\in V}(\alpha_i)_k = \sum_{i\in V}\sum_{k=0}^{n}(\alpha_i)_k \geq \sum_{i\in V}(\beta_i - \varepsilon/\left|V\right|) = \sum_{i\in V}\beta_i - \varepsilon.$$

Hence, $\sum_{n\in\mathbb{N}}\sum_{i\in V}(\alpha_i)_n = \sum_{i\in V}\beta_i$, and the proof is finished. $\square$

We restate the fact that this section is not going to develop a theory of convergence on the real numbers. As such, Proposition 1.69 will be simply stated without proof, and the proof of Theorem 1.70 uses the fact that "the limit of a sum is the sum of limits". We assume such results are no surprise to the reader. We remark that Theorem 1.70, in particular, can be understood in a deeper sense once one understands some basic facts about power series.

Proposition 1.69. Let $\alpha \in \mathbb{R}$ be such that $|\alpha| < 1$. Then

1. $\lim_{n\to\infty} \alpha^n = 0$.
2. $\lim_{n\to\infty} n\alpha^n = 0$

Theorem 1.70. Let $\alpha \in \mathbb{R}$ be such that $|\alpha| < 1$. Then

$$\sum_{n\in\mathbb{N}} n\alpha^{n-1} = \frac{1}{(1-\alpha)^2}.$$

Proof. For every $n \in \mathbb{N}$, define $S_k \coloneqq \sum_{n=0}^{k} n\alpha^n$. Some careful index handling ensures the following equalities:

$$\begin{aligned} S_k &= \sum_{t=2}^{k-1}(t+1)\alpha^t + 1 + 2\alpha \\ \alpha^2 S_k &= \sum_{t=2}^{k-1}(t-1)\alpha^t + (k-1)\alpha^k + k\alpha^{k+1} \\ 2\alpha S_k &= \sum_{t=2}^{k=1} 2t\alpha^t + 2\alpha + 2k\alpha^k \end{aligned}$$

Since $(1-\alpha)^2 = 1+\alpha^2-2\alpha$, we have that

$$(1-\alpha)^2 S_k = 1-\alpha^k(1-\alpha)k-\alpha^k.$$

Therefore

$$\lim_{k\to\infty} S_k = \lim_{k\to\infty} \frac{1-\alpha^k(1-\alpha)k-\alpha^k}{(1-\alpha)^2} = \frac{1}{(1-\alpha)^2} - \frac{1}{(1-\alpha)}\left(\lim_{k\to\infty} k\alpha^k\right) - \frac{1}{(1-\alpha)^2}\left(\lim_{k\to\infty} \alpha^k\right).$$

Proposition 1.69 then finishes the proof. □

1.9 Measure and Probability Theory

A *σ-algebra* on a set X is a collection $\Sigma \subseteq \mathcal{P}(X)$ such that

(i) $\varnothing \in \Sigma$;

(ii) the collection Σ is *closed under complementation*, i.e., if $E \in \Sigma$, then $X \setminus E \in \Sigma$;

(iii) the collection Σ is *closed under countable unions*, that is, if $\mathcal{F} \subseteq \Sigma$ is countable, then

$$\bigcup \mathcal{F} \in \Sigma.$$

For convenience, the set $X \setminus E$ is denoted by E^c when X is clear from the context.

Theorem 1.71. Let X and I be sets, and let $\{\, \Sigma_i : i \in I\}$ be a family of σ-algebras on X. The collection $\bigcap_{i\in I} \Sigma_i$ is a σ-algebra on X.

Proof. Set $\Sigma_I \coloneqq \bigcap_{i\in I} \Sigma_I$. By the definition of σ-algebra, $\varnothing \in \Sigma_i$ for every $i \in I$. Therefore, it also belongs to Σ_I. Suppose then that $E \in \Sigma_I$. Hence, for every $i \in I$, the set E belongs to Σ_i, which implies that $E^c \in \Sigma_i$. Since this holds for every $i \in I$, it follows that $E^c \in \Sigma_I$. Finally, let $\mathcal{F} \subseteq \Sigma_I$ be countable. For every $i \in I$, we have that $\mathcal{F} \subseteq \Sigma_i$, which implies that $\bigcup \mathcal{F} \in \Sigma_i$. Hence, $\bigcup \mathcal{F} \in \Sigma_I$. □

Let X be a set, and let $\mathcal{O} \subseteq \mathcal{P}(X)$. Define the *σ-algebra generated by $\mathcal{O}$* as

$$\sigma(\mathcal{O}) \coloneqq \bigcap \{\, \Sigma \subseteq \mathcal{P}(X) : \Sigma \text{ is a } \sigma\text{-algebra on } X, \mathcal{O} \subseteq \Sigma\}.$$

Theorem 1.71 ensures that $\sigma(\mathcal{O})$ is a σ-algebra. We say that the collection $\mathcal{O}$ *generates* $\sigma(\mathcal{O})$. Note that the power set of X itself is always a σ-algebra on X. Therefore, there is at least one σ-algebra in the intersection when $\sigma(\mathcal{O})$ is being considered. Also, every σ-algebra Σ on X such that $\mathcal{O} \subseteq \Sigma$ will be a superset of $\sigma(\mathcal{O})$. For this reason, $\sigma(\mathcal{O})$ is sometimes refered to as "the smallest σ-algebra containing $\mathcal{O}$".

Theorem 1.72. Let X be a set. Let $\sigma\colon \mathcal{P}(\mathcal{P}(X)) \to \mathcal{P}(\mathcal{P}(X))$ be the function defined, for every $\mathcal{O} \in \mathcal{P}(\mathcal{P}(X))$, as $\sigma(\mathcal{O})$. Then for every $\mathcal{O}$ and $\mathcal{Q}$ in $\mathcal{P}(\mathcal{P}(X))$, we have that

(i) $\mathcal{O} \subseteq \sigma(\mathcal{O})$,

(ii) $\mathcal{Q} \subseteq \mathcal{O} \implies \sigma(\mathcal{Q}) \subseteq \sigma(\mathcal{O})$, and

(iii) If $\mathcal{O}$ is a σ-algebra on X, then $\sigma(\mathcal{O}) = \mathcal{O}$. In particular, $\sigma(\sigma(\mathcal{O})) = \sigma(\mathcal{O})$.

Proof. For every $\mathcal{O} \subseteq \mathcal{P}(X)$, define

$$\Sigma_{\mathcal{O}} \coloneqq \{\, \Sigma : \Sigma \text{ is a } \sigma\text{-algebra}, \mathcal{O} \subseteq \Sigma\},$$

so that $\sigma(\mathcal{O}) = \bigcap \Sigma_{\mathcal{O}}$. By definition, $\mathcal{O} \subseteq \Sigma$ for every $\Sigma \in \Sigma_{\mathcal{O}}$. Hence, $\mathcal{O} \subseteq \bigcap \Sigma_{\mathcal{O}}$, and (i) holds.

If $\mathcal{Q} \subseteq \mathcal{P}(X)$ and $\mathcal{O} \subseteq \mathcal{P}(X)$ are such that $\mathcal{Q} \subseteq \mathcal{O}$, then for every $\Sigma \in \Sigma_{\mathcal{O}}$ we have that $\Sigma \in \Sigma_{\mathcal{Q}}$. In other words, $\Sigma_{\mathcal{O}} \subseteq \Sigma_{\mathcal{Q}}$, which implies (ii).

Item (i) ensures that $\mathcal{O} \subseteq \sigma(\mathcal{O})$. Moreover, if $\mathcal{O}$ is a σ-algebra, then $\mathcal{O} \in \Sigma_{\mathcal{O}}$. Therefore, if $\mathcal{O}$ is a σ-algebra,

$$\sigma(\mathcal{O}) = \bigcap \Sigma_{\mathcal{O}} \subseteq \mathcal{O},$$

so that (iii) holds. □

It is important to contemplate the tool just defined. Given an arbitrary collection $\mathcal{O}$ of subsets of a set X, it is possible to associate it to a σ-algebra on X such that Theorem 1.72 holds. This connection will provide us not only with a tool to define new σ-algebras, but also to conclude things about a given one by looking at a collection that generates it.

A *measurable space* is an ordered pair (X, Σ) where Σ is a σ-algebra on X. Let (X, Σ) be a measurable space. Whenever Σ is clear, the elements $E \in \Sigma$ are simply called *measurable sets of* X.

Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. A function $f \colon X \to Y$ is *measurable* with respect to (w.r.t.) Σ_X and Σ_Y if the preimage of every measurable set in Σ_Y is measurable in Σ_X, that is, if for every $E \in \Sigma_Y$,

$$f^{-1}(E) \in \Sigma_X.$$

Once again, whenever Σ_Y and Σ_X can be inferred from context, they are omitted, and f is simply said to be *measurable.*

Theorem 1.73. Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. Let $\mathcal{O} \in \mathcal{P}(Y)$ be such that $\Sigma_Y = \sigma(\mathcal{O})$. Then a function $f \colon X \to Y$ is measurable if and only if for every $E \in \mathcal{O}$, we have that $f^{-1}(E) \in \Sigma_X$.

Proof. Since $\mathcal{O} \subseteq \Sigma_Y$, if f is measurable then every set in $\mathcal{O}$ is measurable.

Assume then that for every $E \in \mathcal{O}$, we have that $f^{-1}(E) \in \Sigma_X$. Define

$$\Sigma := \{\, E \in \Sigma_Y : f^{-1}(E) \in \Sigma_X \}.$$

Note that the hypothesis ensures that $\mathcal{O} \subseteq \Sigma$. Hence, item (ii) in Theorem 1.72 implies that $\sigma(\mathcal{O}) \subseteq \sigma(\Sigma)$. Moreover, item (iii) in the same theorem reduces the problem in proving that Σ is a σ-algebra, since in such case

$$\Sigma_Y = \sigma(\mathcal{O}) \subseteq \sigma(\Sigma) = \Sigma.$$

We proceed in this direction. First note that $\varnothing \in \Sigma$, since $\varnothing \in \Sigma_Y$ and $f^{-1}(\varnothing) = \varnothing \in \Sigma_X$. Suppose $E \in \Sigma$. Then $E \in \Sigma_Y$ and, therefore, $E^c \in \Sigma_Y$. Moreover, Proposition 1.1 ensures $f^{-1}(E^c) = f^{-1}(E)^c$, so that $E^c \in \Sigma$. Finally, let $\mathcal{F} \subseteq \Sigma$ be countable. By definition, we have that $\bigcup \mathcal{F} \in \Sigma_Y$. Proposition 1.1 ensures that $f^{-1}(\bigcup \mathcal{F}) = \bigcup_{F \in \mathcal{F}} f^{-1}(F)$. Hence, $\bigcup \mathcal{F} \in \Sigma$, and the proof is done. □

Definition 1.74. Let (X, Σ) be a measurable space. A function $\mu \colon X \to \mathbb{R}_+ \cup \{\infty\}$ is a *measure* on (X, Σ) if

(i) $\mu(\varnothing) = 0$,

(ii) μ is *countably additive*, that is, if $(E_i)_{i \in \mathbb{N}}$ is a sequence of pairwise disjoint elements in Σ, then

$$\mu\left(\bigcup_{i \in \mathbb{N}} E_i\right) = \sum_{i \in \mathbb{N}} \mu(E_i).$$

When μ is a measure on the measurable space (X, Σ), we say that (X, Σ, μ) is a *measure space.*

Let (X, Σ_X, μ) be a measure space, and (Y, Σ_Y) be a measurable space. A measurable function $f \colon X \to Y$ hints at a measure for Y. For every $E \in \Sigma_Y$, define $\nu \colon \Sigma_Y \to \mathbb{R}_+ \cup \{\infty\}$ as

$$\nu(E) := \mu(f^{-1}(E)).$$

Note that the fact that f is measurable ensures that μ is defined on every set of the form $f^{-1}(E)$, with $E \in \Sigma_Y$. This and Proposition 1.1 ensure that ν is a measure on (Y, Σ_y). This construction will play a central role in the arguments on section 3.3, most notably in Proposition 3.19 and in Proposition 3.28.

We now focus on the probability theory concepts that will be used in the text. A measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space* if $\mathbb{P}(\Omega) = 1$. In this case, $\mathbb{P}$ is said to be a *probability measure* or a *probability distribution.* Whenever working with a probability space, the elements $E \in \mathcal{F}$ are called *events.* Also, a *random variable* is any measurable function whose domain is a probability space.

Let S and T be sets, and let $f \colon S \to T$ be any function. For every $U \subseteq T$, define

$$\{f \in U\}_S := \{\, s \in S : f(s) \in U \}.$$

This notation will be used throughout the text, mostly when working with measurable functions. To understand why, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(V, \mathcal{G})$ be a measure space, and let $X\colon \Omega \to V$ be a random variable. We have that, for every measurable set $E \in \mathcal{G}$, the set

$$X^{-1}(E) = \{\, \omega \in \Omega : X(\omega) \in E\} = \{X \in E\}_{\Omega}$$

is an element of $\mathcal{F}$, i.e., an event. For this reason, this notation will be abundant in our demonstrations. Furthermore, for any events $F, G \in \mathcal{F}$, we define

$$\mathbb{P}(F, G) \coloneqq \mathbb{P}(F \cap G).$$

Note that the above notations can be combined. If, for example, $(W, \mathcal{H})$ is any measurable space — maybe even $(V, \mathcal{G})$ again — and $Y\colon \Omega \to W$ is a random variable, we have that for any $E \in \mathcal{G}$ and $F \in \mathcal{H}$,

$$\mathbb{P}(X \in E, Y \in F) = \mathbb{P}(\{X \in E, y \in F\}_{\Omega}) = \mathbb{P}(\{X \in E\}_{\Omega} \cap \{Y \in F\}_{\Omega}).$$

Whenever V is a finite or countable set, we simply reffer to V as a measurable space, and have, by definition, $\mathcal{P}(V)$ as its σ-algebra. In other words, if V is at most countable, then the measurable space $(V, \mathcal{P}(V))$ will be reffered to as the measurable space V. Note that, in this case,

$$\sigma(\{\, \{i\} : i \in V\}) = \mathcal{P}(V).$$

Hence, if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then Theorem 1.73 implies that to prove that a function $X\colon \Omega \to V$ is a random variable suffices to prove that for every $i \in V$ we have that $\{X = i\}_{\Omega}$ is an event.

On a similar note, when working with the set of real numbers $\mathbb{R}$, the σ-algebra defined in it will always be

$$\sigma(\{\, (\alpha, \beta) \subseteq \mathbb{R} : \alpha \in \mathbb{R}, \beta \in \mathbb{R}\}).$$

Therefore, if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and we say that a function $X\colon \Omega \to \mathbb{R}$ is measurable, we are saying that for every $\alpha, \beta \in \mathbb{R}$, the set

$$\{X \in (\alpha, \beta)\}_{\Omega} = \{\alpha < X, X < \beta\}_{\Omega}$$

is measurable.

Definition 1.75. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X\colon \Omega \to \mathbb{R}$ be a random variable. Suppose that there exists a countable set $I \subseteq \mathbb{R}$ such that

$$\sum_{x \in I} \mathbb{P}(X = x) = 1.$$

Then the *expected value of* X is defined as

$$\mathbb{E}[X] \coloneqq \sum_{x \in I} x\mathbb{P}(X = x).$$

It is possible to define the expected value of any random variable. This is actually the main goal of measure theory, and is done by defining the integral of a measurable function. However, for the purposes of this work, the above definition will be much more convenient, and, for this reason, will be the one adopted.

Definition 1.76. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $E \in \mathcal{F}$ be an event such that $\mathbb{P}(E) > 0$. For every event $A \in \mathcal{F}$, the *conditional probability of* A *given* E is defined as

$$\mathbb{P}(A \mid E) \coloneqq \frac{\mathbb{P}(A, E)}{\mathbb{P}(E)}.$$

Note that $(\Omega, \mathcal{F}, \mathbb{P}(\cdot \mid E))$ is also a probability space.

The definition of conditional probability provides a natural definition of conditional expectation. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X\colon \Omega \to \mathbb{R}$ is a random variable. Suppose that there exists a countable set $I \subseteq \mathbb{R}$ such that $\sum_{x\in I} \mathbb{P}(X = x) = 1$. Let $E \in \mathcal{F}$ be any event such that $\mathbb{P}(E) > 0$. Then the *conditional expectation of X with respect to E* is defined as

$$\mathbb{E}[X \mid E] \coloneqq \sum_{x\in I} x\mathbb{P}(X = x \mid E).$$

Another simple way to extend the idea of conditional probability is to "nest" it. Suppose then that $F \in \mathcal{F}$ is such that $\mathbb{P}(E, F) > 0$. Note that this implies that $\mathbb{P}(F) > 0$. Then for every $A \in \mathcal{F}$,

$$\mathbb{P}(A \mid E, F) = \frac{\mathbb{P}(A, E, F)}{\mathbb{P}(E, F)} = \frac{\mathbb{P}(A, E \mid F)\mathbb{P}(F)}{\mathbb{P}(E \mid F)\mathbb{P}(F)} = \frac{\mathbb{P}(A, E \mid F)}{\mathbb{P}(E \mid F)}.$$

Hence, to condition on F, and use this new probability measure to condition on E is equivalent to conditioning on the event $E \cap F$.

Definition 1.77. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A set of events $\mathcal{I} \subseteq \mathcal{F}$ is said to be *independent* if for every finite subset $\mathcal{S}$ of it, it holds that

$$\mathbb{P}\left(\bigcap \mathcal{S}\right) = \prod \mathbb{P}(\mathcal{S}).$$

Definition 1.78. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let I be a set, and $\{\, (M_t, \Sigma_t) : t \in I\}$ be a collection of measurable spaces. A set of random variables $\{\, X_t \in M_t^{\Omega} : t \in I\}$ is said to be *independent* if for every finite subset $\mathcal{S}$ of it and for every collection of events $\{\, E_t \in \Sigma_t : t \in \mathcal{S}\}$,

$$\mathbb{P}\left(\bigcap_{t\in\mathcal{S}} \{X_t = E_t\}_{\Omega}\right) = \prod_{t\in\mathcal{S}} \mathbb{P}(X_t = E_t).$$

Chapter 2

The Naive Algorithm

The naive algorithm relies on Tutte's Matrix Tree Theorem, which relies on the Cauchy-Binet Formula. This is the roadmap for this chapter. The naive algorithm was first described by Kulkarni [18], in a more general setting, more focused on sampling than on spanning trees. The approach here is quite different, focusing on the underlying theory — most importantly, on Kirchhoff's Matrix Tree Theorem [17] and Tutte's generalization.

It is interesting to note that the algorithm developed in this chapter will actually sample r-arborescences. We begin then by showing this is equivalent to the problem of sampling spanning trees.

Proposition 2.1. For every $T \in \mathcal{T}_G$ and for every $r \in V$, there is a unique $F \in \mathcal{T}_D(r)$ such that the underlying graph of $D[F]$ is $G[T]$.

Proof. This can be done by induction on $|V|$.

If $|V| = 0$, the thesis vacuously holds. If $|V| = 1$, then both $\mathcal{T}_G$ and $\mathcal{T}_D(r)$ are equal to $\{\varnothing\}$.

Let then $|V| > 1$. Let $T \in \mathcal{T}_G$ and let j be a leaf of $G[T]$ which is distinct from r. Let k be the only vertex adjacent to j in $G[T]$. Note that $T \setminus \{kj\}$ is a spanning tree of $G - j$. Therefore, the induction hypothesis applies, and there is a unique $F' \in \mathcal{T}_{D-j}(r)$ such that $(G-j)[T \setminus \{kj\}]$ is the underlying graph of $(D-j)[F']$.

Define $F := \{kj\} \cup F'$. Note then that $D[F]$ is an r-arborescence of D whose underlying graph is $G[T]$. Note that, since $r \neq j$, then we have that $kj \in F$ in every r-arborescence F of D, to ensure that $\left|\delta^{in}(j)\right| = 1$. This, along with the fact that F' is unique, implies the uniqueness of F. □

2.1 The Cauchy-Binet Formula

The main issue we address here is the fact that the determinant is defined on matrices with the same set of row and column indices. We need to extend such definition, not only to apply it to incidence matrices, but even to work with the determinant of submatrices. Bourbaki [4] defines matrices as we do, namely, as a function whose domain is the product of two finite sets. We work with his definition, and use the idea of function matrix, to define the determinant of a matrix when the set of row indices and column indices is not the same, but of the same size.

Such definitions and developments are new. They arose from necessity; even to just state the Cauchy-Binet Formula precisely, it is necessary to develop new concepts. A function matrix is a meaningful and precise way of relating the new concepts that must be approached with the definitions already present in Chapter 1.

Lemma 2.2. Let U and V be finite sets. Let $A\colon U \times V \to \mathbb{R}$ be any function. Then

$$\prod_{i \in U} \sum_{j \in V} A(i,j) = \sum_{f\colon U \to V} \prod_{i \in U} A(i, f(i)).$$

Proof. This proof can be done by induction on $|U|$.

If $|U| = 0$, then $U = \varnothing$. It is a (curious) vacuous truth that there is a unique function $f\colon \varnothing \to V$. Denying either its existence or its uniqueness require an element in the empty set. Therefore, the RHS is the sum of only one product, and this product is empty. Since the LHS is also an empty product, it follows that both sides are equal to 1, and the base case holds.

Suppose that $|U| > 0$. Take any $k \in U$, and set $U' \coloneqq U \setminus \{k\}$. Since for any function $f\colon U \to V$ it is true that

$$1 = \sum_{j \in V} [f(k) = j],$$

we can multiply the summation over functions by 1, factor the term with k, and obtain that

$$\begin{aligned}\sum_{f\colon U \to V} \prod_{i \in U} A(i, f(i)) &= \sum_{f\colon U \to V} \left(\sum_{j \in V} [f(k) = j] \right) \prod_{i \in U} A(i, f(i)) \\ &= \sum_{j \in V} A(k, j) \left(\sum_{f\colon U \to V} [f(k) = j] \prod_{i \in U'} A(i, f(i)) \right).\end{aligned}$$

This restricts the sum over all the functions $g\colon U' \to V$, and the induction hypothesis completes the proof:

$$\sum_{j \in V} A(k, j) \left(\sum_{g\colon U' \to V} \prod_{i \in U'} A(i, g(i)) \right) = \left(\sum_{j \in V} A(k, j) \right) \left(\prod_{i \in U'} \sum_{j \in V} A(i, j) \right) = \prod_{i \in U} \sum_{j \in V} A(i, j). \qquad \square$$

Definition 2.3. Let U and V be finite sets. Let $f\colon U \to V$ be a function. The *function matrix* $P_f \in \mathbb{R}^{V \times U}$ is defined as

$$P_f \coloneqq \sum_{i \in U} e_{f(i)} e_i^{\mathsf{T}}.$$

The usefulness of a function matrix comes from the fact that it is an algebraic object that encodes the operation of applying a function to the indices of a matrix. More formally, let U and V be finite sets, and $f\colon U \to V$ be a function. For any $i \in U$,

$$P_f e_i = \left(\sum_{j \in U} e_{f(j)} e_j^{\mathsf{T}} \right) e_i = \sum_{j \in U} [i = j] e_{f(j)} = e_{f(i)}.$$

We have already worked with an example of function matrix, and obtained several interesting results. Let $D = (V, A, \psi)$ be a digraph. Moreover, let $\lambda\colon V \times V \to V$ and $\rho\colon V \times V \to V$ be defined as, for every $i, j \in V$,

$$\begin{aligned}\lambda(i, j) &\coloneqq i, \\ \rho(i, j) &\coloneqq j.\end{aligned}$$

Furthermore, let $a \in A$ be such that $\psi(a) = ij$. Then

$$H_D e_a = e_j = e_{\rho\psi(a)} = P_{\rho\psi} e_a.$$

Since this argument holds for every $a \in A$, we conclude that $H_D = P_{\rho\psi}$. Similarly, we have that $T_D = P_{\lambda\psi}$. Hence, the computations involving the head and tail matrices are an example of how convenient it is to compute with function matrices, since the matrix product reduces into function application.

Proposition 2.4. Let T, U and V be finite sets. Let $f\colon U \to V$ and $g\colon T \to U$. Then

$$P_f P_g = P_{fg}.$$

Proof. Note that, for every $i \in T$,

$$P_f P_g e_i = P_f e_{g(i)} = e_{fg(i)} = P_{fg} e_i.$$

Since the set $\{e_i \in \mathbb{R}^T : i \in T\}$ generates $\mathbb{R}^V$, this suffices to prove the desired equation. $\square$

Proposition 2.5. Let U and V be finite sets. Let $f\colon U \to V$ be a bijective function. Then

$$P_{f^{-1}} = P_f^{\mathsf{T}}.$$

Proof. For every $i \in V$, note that

$$\begin{aligned}(P_f)^{\mathsf{T}} e_i &= \left(\sum_{j\in U} e_{f(j)} e_j^{\mathsf{T}}\right)^{\mathsf{T}} e_i \\ &= \sum_{j\in U} [f(j) = i] e_j \\ &= e_{f^{-1}(i)} = P_{f^{-1}} e_i.\end{aligned}$$

Since the set $\{e_i \in \mathbb{R}^V : i \in V\}$ generates $\mathbb{R}^V$, this completes the proof. □

Function matrices represent a simple linear transformation between vector spaces, which uses the given function to associate elements from the canonical basis. As a result, it is possible to simplify the products quite easily. Let $A \in \mathbb{R}^{V\times U}$ and $\phi\colon V \to U$. Then

$$(AP_\phi)_{i,j} = e_i^{\mathsf{T}} A P_\phi e_j = e_i^{\mathsf{T}} A e_{\phi(j)} = A_{i,\phi(j)}. \tag{2.6}$$

Moreover, if ϕ is bijective,

$$(P_\phi A)_{i,j} = e_i^{\mathsf{T}} P_\phi A e_j = (P_\phi^{\mathsf{T}} e_i)^{\mathsf{T}} A e_j = A_{\phi^{-1}(i),j}. \tag{2.7}$$

Note that given a matrix $A \in \mathbb{R}^{V\times U}$ and a bijective function $\phi\colon V \to U$, there are actually two ways to have a matrix with the same row and column set — AP_ϕ and $P_\phi A$. The former describes an operator on $\mathbb{R}^V$, and the latter an operator on $\mathbb{R}^U$. It is then possible to calculate both determinants, both of which somehow compete for the (yet to come) definition of "determinant of A with respect to ϕ". The next proposition will avoid such a crisis, by ensuring that both calculations lead to the same result.

Proposition 2.8. Let U and V be finite sets. Let $\phi\colon V \to U$ be a bijective function. If $A \in \mathbb{R}^{V\times U}$, then

$$\det(AP_\phi) = \det(P_\phi A).$$

Proof. Lemma 1.20 implies that for any $\sigma \in \operatorname{Sym}(V)$,

$$\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma\phi^{-1}\phi) = \operatorname{sgn}(\phi\sigma\phi^{-1}).$$

Therefore, equation (2.6) ensures that

$$\begin{aligned}\det(AP_\phi) &= \sum_{\sigma\in\operatorname{Sym}(V)} \operatorname{sgn}(\sigma) \prod_{i\in V} A_{i,\phi\sigma(i)} \\ &= \sum_{\sigma\in\operatorname{Sym}(V)} \operatorname{sgn}(\phi\sigma\phi^{-1}) \prod_{i\in V} A_{i,\phi\sigma(i)}.\end{aligned}$$

Since the mapping $(\sigma \mapsto \phi\sigma\phi^{-1})$ is a bijection from $\operatorname{Sym}(V)$ to $\operatorname{Sym}(U)$, we can change the summation index and apply equation (2.7), to conclude that

$$\begin{aligned}\det(AP_\phi) &= \sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\tau) \prod_{i\in V} A_{i,\tau\phi(i)} \\ &= \sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\tau) \prod_{i\in U} A_{\phi^{-1}(i),\tau(i)} \\ &= \sum_{\tau\in\operatorname{Sym}(U)} \operatorname{sgn}(\tau) \prod_{i\in U} (P_\phi A)_{i,\tau(i)} \\ &= \det(P_\phi A).\end{aligned}$$

□

The proof above demands a remark. Let $f, g \colon U \to V$ be bijective functions. Lemma 1.20 implies that

$$\operatorname{sgn}(gf^{-1}) = \operatorname{sgn}(f^{-1}g).$$

The LHS is the sign of a permutation on V, and the RHS is the sign of a permutation on U. Proposition 2.8 translates this result to a different concept, since for $A \in \mathbb{R}^{V \times U}$ and $\phi \colon V \to U$ bijective, we just proved that

$$\det(AP_\phi) = \det(P_\phi A).$$

The LHS is the determinant of a matrix on $\mathbb{R}^{V \times V}$, and the RHS is the determinant of a matrix on $\mathbb{R}^{U \times U}$.

Definition 2.9. Let U and V be finite sets. Let $\phi \colon V \to U$ be a bijective function. Let $A \in \mathbb{R}^{V \times U}$. The *determinant* (with respect to ϕ) of A is defined as

$$\det_\phi(A) \coloneqq \det(AP_\phi).$$

Theorem 2.10. Let U and V be finite sets. Let $\phi \colon V \to U$ be a bijective function. If $A \in \mathbb{R}^{V \times U}$, then

$$\det_\phi(A) = \det_{\phi^{-1}}(A^\mathsf{T}).$$

Proof. Several previous results come into play. Applying successively, Theorem 1.23, Proposition 2.5, and Proposition 2.8, we have that

$$\det_\phi(A) = \det(AP_\phi) = \det(P_\phi^\mathsf{T} A^\mathsf{T}) = \det(P_{\phi^{-1}} A^\mathsf{T}) = \det(A^\mathsf{T} P_{\phi^{-1}}) = \det_{\phi^{-1}}(A^\mathsf{T}). \qquad \square$$

Theorem 2.11. Let U and V be finite sets. Let $f \colon U \to V$ and $g \colon U \to V$ be functions. Then

$$\det(P_f^\mathsf{T} P_g) = [f, g \text{ injective}][\operatorname{Im}(f) = \operatorname{Im}(g)] \operatorname{sgn}(f^{-1}g).$$

Proof. We will prove:

(1) If $\det(P_f^\mathsf{T} P_g)$ is nonzero, then both f and g are injective.

(2) If $\det(P_f^\mathsf{T} P_g)$ is nonzero, then $\operatorname{Im}(f) = \operatorname{Im}(g)$.

(3) If $\det(P_f^\mathsf{T} P_g)$ is nonzero, then it is equal to $\operatorname{sgn}(f^{-1}g)$.

First, note that if f is not injective, then for any $A \in \mathbb{R}^{V \times U}$, we have that $\det(P_f^\mathsf{T} A) = 0$. To see why, assume there are distinct i and j in U such that $f(i) = f(j)$. Then

$$P_f e_i = e_{f(i)} = e_{f(j)} = P_f e_j,$$

so that $P_f(e_i - e_j) = 0$. It follows that $e_i - e_j$ is a nonzero vector in $\operatorname{Null}(P_f)$, and, therefore, in $\operatorname{Null}(A^\mathsf{T} P_f)$. This implies that $\det(A^\mathsf{T} P_f) = 0$, and Theorem 1.23 ensures that $\det(P_f^\mathsf{T} A) = 0$.

The contrapositive of this result applied to both $P_f^\mathsf{T} P_g$ and $P_g^\mathsf{T} P_f$ implies (1).

Now the second step. Assume that $\operatorname{Im}(g) \not\subseteq \operatorname{Im}(f)$. Then there exists $i \in U$ such that for every $j \in U$ we have that $f(j) \neq g(i)$. Therefore, for every $j \in U$,

$$0 = e_{f(j)}^\mathsf{T} e_{g(i)} = e_j^\mathsf{T} P_f^\mathsf{T} P_g e_i.$$

In other words, $e_i \in \operatorname{Null}(P_f^\mathsf{T} P_g)$, so that $\det(P_f^\mathsf{T} P_g)$ must be zero. Therefore, $\operatorname{Im}(g) \not\subseteq \operatorname{Im}(f)$ implies that $\det(P_f^\mathsf{T} P_g)$ is zero.

The contrapositive of this result applied to both $P_f^\mathsf{T} P_g$ and $P_g^\mathsf{T} P_f$ implies step (2).

Now the final step. Let $f, g \colon U \to V$ be such that $\det(P_f^\mathsf{T} P_g)$ is nonzero. Results (1) and (2) imply that there is $S \subseteq V$ such that $\operatorname{Im}(f) = \operatorname{Im}(g) = S$ and that there exists $g^{-1} \colon S \to U$, inverse of g.

For every $i, j \in U$,

$$(P_f^\mathsf{T} P_g)_{ij} = e_i^\mathsf{T} P_f^\mathsf{T} P_g e_j = e_{f(i)}^\mathsf{T} e_{g(j)} = [f(i) = g(j)].$$

Hence,

$$\begin{aligned}\det(P_f^{\mathsf{T}} P_g) &= \sum_{\sigma \in \mathrm{Sym}(U)} \operatorname{sgn}(\sigma) \prod_{i \in U} (P_f^{\mathsf{T}} P_g)_{i,\sigma(i)} \\ &= \sum_{\sigma \in \mathrm{Sym}(U)} \operatorname{sgn}(\sigma) \prod_{i \in U} [f(i) = g\sigma(i)] \\ &= \sum_{\sigma \in \mathrm{Sym}(U)} \operatorname{sgn}(\sigma) [f = g\sigma].\end{aligned}$$

Given the conditions on f and g, it holds that $f = g\sigma$ if and only if $\sigma = g^{-1} f$, so that

$$\begin{aligned}\det(P_f^{\mathsf{T}} P_g) &= [f, g \text{ injective}][\mathrm{Im}(f) = \mathrm{Im}(g)] \sum_{\sigma \in \mathrm{Sym}(U)} \operatorname{sgn}(\sigma)[\sigma = g^{-1} f] \\ &= [f, g \text{ injective}][\mathrm{Im}(f) = \mathrm{Im}(g)] \operatorname{sgn}(g^{-1} f).\end{aligned}$$

The proof is complete since, if f and g are injective and have the same image, the function $g^{-1} f$ is invertible and its inverse is $f^{-1} g$, so that $\operatorname{sgn}(g^{-1} f) = \operatorname{sgn}(f^{-1} g)$. □

The theorem just proved goes further into the direction of relating determinants of matrices and signs of permutations. Let $f, g \colon U \to V$ be functions. Note that Proposition 2.5 hints that P_f^{T} is a "substitute" for f^{-1}, and the result just proved says that $\det(P_f^{\mathsf{T}} P_g)$ is a good generalization for $\operatorname{sgn}(f^{-1} g)$, since both are equal whenever the expression $\operatorname{sgn}(f^{-1} g)$ makes sense, i.e., $f^{-1} g$ exists and is invertible.

Proposition 2.12. Let U and V be finite sets. Let $S \subseteq V$. Let $f \colon U \to V$. Let $\phi \colon S \to U$ be a bijective function. Then

$$\det_{\phi}(P_f[S, U]) = [f \text{ injective}][\mathrm{Im}(f) = S] \operatorname{sgn}(\phi f).$$

Proof. Proposition 2.5 and Proposition 2.8 ensure that

$$\det_{\phi}(P_f[S, U]) = \det(P_f[S, U] P_\phi) = \det(P_\phi P_f[S, U]) = \det(P_{\phi^{-1}}^{\mathsf{T}} P_f[S, U]).$$

Moreover, note that

$$P_f[S, U] e_i = [f(i) \in S] e_{f(i)}.$$

Hence, if $\mathrm{Im}(f) \not\subseteq S$, we have that there exists $i \in U$ such that $P_f[S, U] e_i = 0$. This implies that both sides of the statement are zero. We can then assume that $\mathrm{Im}(f) \subseteq S$. In this case, we have that $P_f[S, U]$ is a function matrix in $\mathbb{R}^{S \times U}$. Furthermore, ϕ is bijective, $\mathrm{Im}(\phi^{-1}) = S$, so that Theorem 2.11 ensures

$$\det_{\phi}(P_f[S, U]) = [f \text{ injective}][\mathrm{Im}(f) = S] \operatorname{sgn}(\phi f).$$ □

Proposition 2.13 (Cauchy-Binet, restricted version)**.** Let U and V be finite sets. Let $f, g \colon U \to V$ be functions. For every set $S \in \binom{V}{|U|}$, let $\phi_S \colon S \to U$ be a bijective function. Then

$$\det(P_f^{\mathsf{T}} P_g) = \sum_{S \in \binom{V}{|U|}} \det_{\phi_S^{-1}}(P_f^{\mathsf{T}}[U, S]) \det_{\phi_S}(P_g[S, U]).$$

Proof. Theorem 2.10 and Proposition 2.12 ensure that

$$\begin{aligned}\sum_{S \in \binom{V}{|U|}} \det_{\phi_S^{-1}}(P_f^{\mathsf{T}}[U, S]) \det_{\phi_S}(P_g[S, U]) &= \sum_{S \in \binom{V}{|U|}} \det_{\phi_S}(P_f[S, U]) \det_{\phi_S}(P_g[S, U]) \\ &= \sum_{S \in \binom{V}{|U|}} [f \text{ injective}][\mathrm{Im}(f) = S] \operatorname{sgn}(\phi_S f)[g \text{ injective}][\mathrm{Im}(g) = S] \operatorname{sgn}(\phi_S g) \\ &= [f, g \text{ injective}][\mathrm{Im}(f) = \mathrm{Im}(g)] \sum_{S \in \binom{V}{|U|}} [\mathrm{Im}(f) = S] \operatorname{sgn}(\phi_S f) \operatorname{sgn}(\phi_S g) \\ &= [f, g \text{ injective}][\mathrm{Im}(f) = \mathrm{Im}(g)] \operatorname{sgn}(\phi_{\mathrm{Im}(f)} f) \operatorname{sgn}(\phi_{\mathrm{Im}(g)} g).\end{aligned}$$

Let $S \coloneqq \operatorname{Im}(f)$. If f is injective, we have that $\operatorname{sgn}(\phi_S f) = \operatorname{sgn}(f^{-1}\phi_S^{-1})$, and it is possible to simplify the expression on the nonzero case to

$$\operatorname{sgn}(\phi_S g)\operatorname{sgn}(\phi_S f) = \operatorname{sgn}(f^{-1}\phi_S^{-1})\operatorname{sgn}(\phi_S g) = \operatorname{sgn}(f^{-1}g).$$

Note that Theorem 2.11 finishes the proof:

$$\sum_{S\in\binom{V}{|U|}} \det_{\phi_S^{-1}}(P_f^{\mathsf{T}}[U,S])\det_{\phi_S}(P_g[S,U]) = [f,g \text{ injective}][\operatorname{Im}(f)=\operatorname{Im}(g)]\operatorname{sgn}(f^{-1}g) = \det(P_f^{\mathsf{T}}P_g). \qquad \square$$

For given functions $f, g\colon U \to V$, the summation on the statement of Proposition 2.13 is precisely to "try all" candidates for $\operatorname{Im}(f)$ and $\operatorname{Im}(g)$. This will generalize into the Cauchy-Binet Formula, but it remains to relate the determinant of arbitrary matrices with the determinant of function matrices.

Proposition 2.14. Let U and V be finite sets. Let $A, B \in \mathbb{R}^{V\times U}$. Then

$$\det(A^{\mathsf{T}}B) = \sum_{f\colon U\to V}\det(P_f^{\mathsf{T}}B)\prod_{i\in U}A_{f(i),i}.$$

Proof. After applying the definition of the determinant, matrix product, and transpose, we obtain

$$\begin{aligned}\det(A^{\mathsf{T}}B) &= \sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\prod_{i\in U}(A^{\mathsf{T}}B)_{i,\sigma(i)}\\ &= \sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\prod_{i\in U}\sum_{j\in V}A^{\mathsf{T}}_{i,j}B_{j,\sigma(i)}\\ &= \sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\prod_{i\in U}\sum_{j\in V}A_{j,i}B_{j,\sigma(i)}.\end{aligned}$$

Now Lemma 2.2 produces the summation over functions needed. Then some factoring and collecting finishes the proof:

$$\begin{aligned}\det(A^{\mathsf{T}}B) &= \sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\sum_{f\colon U\to V}\prod_{i\in U}A_{f(i),i}B_{f(i),\sigma(i)}\\ &= \sum_{f\colon U\to V}\left(\prod_{i\in U}A_{f(i),i}\right)\sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\prod_{i\in U}B_{f(i),\sigma(i)}\\ &= \sum_{f\colon U\to V}\left(\prod_{i\in U}A_{f(i),i}\right)\sum_{\sigma\in\operatorname{Sym}(U)}\operatorname{sgn}(\sigma)\prod_{i\in U}(P_f^{\mathsf{T}}B)_{i,\sigma(i)}\\ &= \sum_{f\colon U\to V}\left(\prod_{i\in U}A_{f(i),i}\right)\det(P_f^{\mathsf{T}}B). \qquad \square\end{aligned}$$

Corollary 2.15. Let U and V be finite sets. Let $A, B \in \mathbb{R}^{V\times U}$. Then

$$\det(A^{\mathsf{T}}B) = \sum_{f\colon U\to V}\sum_{g\colon U\to V}\left(\prod_{i\in U}A_{f(i),i}\right)\left(\prod_{i\in U}B_{g(i),i}\right)\det(P_f^{\mathsf{T}}P_g).$$

Proof. Apply, in this order, Proposition 2.14, Theorem 1.23, Proposition 2.14, and Theorem 1.23.

$$\begin{aligned}
\det(A^{\mathsf{T}}B) &= \sum_{f\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\det(P_f^{\mathsf{T}}B)\\
&= \sum_{f\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\det(B^{\mathsf{T}}P_f)\\
&= \sum_{f\colon U\to V}\sum_{g\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\left(\prod_{i\in U} B_{g(i),i}\right)\det(P_g^{\mathsf{T}}P_f)\\
&= \sum_{f\colon U\to V}\sum_{g\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\left(\prod_{i\in U} B_{g(i),i}\right)\det(P_f^{\mathsf{T}}P_g). \qquad \square
\end{aligned}$$

Theorem 2.16 (The Cauchy-Binet Formula)**.** Let U and V be finite sets. For every set $S \in \binom{V}{|U|}$, let $\phi_S\colon S \to U$ be a bijective function. Let $A, B \in \mathbb{R}^{V\times U}$. Then

$$\det(A^{\mathsf{T}}B) = \sum_{S\in\binom{V}{|U|}} \det_{\phi_S^{-1}}(A^{\mathsf{T}}[U,S])\det_{\phi_S}(B[S,U]).$$

Proof. First, it is useful to give an alternate expression for $\det_{\phi_S}(B[S,U])$.

Let $S \in \binom{V}{|U|}$. Theorem 1.23 and Proposition 2.8 allow the manipulations of the matrices, and Proposition 2.5 states the relation between P_{ϕ_S} and its transpose, so that

$$\det_{\phi_S}(B[S,U]) = \det(B[S,U]P_{\phi_S}) = \det(P_{\phi_S^{-1}}B^{\mathsf{T}}[U,S]) = \det(B^{\mathsf{T}}[U,S]P_{\phi_S^{-1}}).$$

Corollary 2.15 provides the sum over functions, and once again Theorem 1.23, Proposition 2.8, and Proposition 2.5 simplify the determinant, so that

$$\det_{\phi_S}(B[S,U]) = \det(B^{\mathsf{T}}[U,S]P_{\phi_S^{-1}}) = \sum_{g\colon U\to S}\left(\prod_{i\in U} B_{g(i),i}\right)\det(P_g^{\mathsf{T}}P_{\phi_S^{-1}}) = \sum_{g\colon U\to S}\left(\prod_{i\in U} B_{g(i),i}\right)\det_{\phi_S}(P_g).$$

Finally, Proposition 2.12 ensures that the summation range can be extended over every function $g\colon U \to V$, with $\det_{\phi_S}(P_g[S,U])$ selecting the ones whose image is S, so that

$$\det_{\phi_S}(B[S,U]) = \sum_{g\colon U\to V}\left(\prod_{i\in U} B_{g(i),i}\right)\det_{\phi_S}(P_g[S,U]).$$

The path here is clear. Use Corollary 2.15 to write the product in terms of function matrices, then use Proposition 2.13 and the equality just proved to finish the proof:

$$\begin{aligned}
\det(A^{\mathsf{T}}B) &= \sum_{f\colon U\to V}\sum_{g\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\left(\prod_{i\in U} B_{g(i),i}\right)\det(P_f^{\mathsf{T}}P_g)\\
&= \sum_{f\colon U\to V}\sum_{g\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\left(\prod_{i\in U} B_{g(i),i}\right)\sum_{S\in\binom{V}{|U|}}\det_{\phi_S^{-1}}(P_f^{\mathsf{T}}[U,S])\det_{\phi_S}(P_g[S,U])\\
&= \sum_{S\in\binom{V}{|U|}}\sum_{f\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\det_{\phi_S}(P_f[S,U])\sum_{g\colon U\to V}\left(\prod_{i\in U} B_{g(i),i}\right)\det_{\phi_S}(P_g[S,U])\\
&= \sum_{S\in\binom{V}{|U|}}\sum_{f\colon U\to V}\left(\prod_{i\in U} A_{f(i),i}\right)\det_{\phi_S}(P_f[S,U])\det_{\phi_S}(B[S,U])\\
&= \sum_{S\in\binom{V}{|U|}}\det_{\phi_S}(A[S,U])\det_{\phi_S}(B[S,U]) = \sum_{S\in\binom{V}{|U|}}\det_{\phi_S^{-1}}(A^{\mathsf{T}}[U,S])\det_{\phi_S}(B[S,U]). \qquad \square
\end{aligned}$$

From a computational perspective, Theorem 2.16 is interesting because it reduces the sum of an exponential amount of determinants into a single one. This will be used first to give a determinant formula for counting spanning trees of a graph. Then, since determinants can be calculated in polynomial time, this will be fundamental to describe the first of the two algorithms in this text for sampling spanning trees.

2.2 Calculating Determinants

The results developed so far can be exploited further into tools to calculate determinants. The main results on this section are the Matrix Determinant Lemma 2.23 and the Laplace Expansion 2.21. However, the next section will mainly use Proposition 2.20, which can be seen as a "weaker version" of the Laplace Expansion. We begin with a remarkable first application of the Cauchy-Binet Formula.

Proposition 2.17. Let U be a finite set, and $A, B \in \mathbb{R}^{U \times U}$. Then

$$\det(AB) = \det(A)\det(B).$$

Proof. Note that $\mathcal{S} \coloneqq \binom{U}{|U|}$ is actually equal to $\{U\}$. Let $\phi\colon U \to U$ denote the identity function in U. Then Theorem 2.16 ensures

$$\det(AB) = \underset{\phi^{-1}}{\det}(A[U,U])\,\underset{\phi}{\det}(B[U,U]) = \underset{\phi^{-1}}{\det}(A)\,\underset{\phi}{\det}(B) = \det(AI)\det(BI) = \det(A)\det(B). \qquad \square$$

Proposition 2.18. Let U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$. For every $\sigma \in \mathrm{Sym}(U)$ and for every bijection $\phi\colon V \to U$,

$$\underset{\sigma\phi}{\det}(A) = \mathrm{sgn}(\sigma)\,\underset{\phi}{\det}(A).$$

Proof. Apply Proposition 2.4, Proposition 2.8, Proposition 1.25, and Proposition 2.8 again

$$\underset{\sigma\phi}{\det}(A) = \det(AP_{\sigma\phi}) = \det(AP_\sigma P_\phi) = \det(P_\phi A P_\sigma) = \mathrm{sgn}(\sigma)\det(P_\phi A) = \mathrm{sgn}(\sigma)\,\underset{\phi}{\det}(A). \qquad \square$$

Lemma 2.19. Let U be a finite set, and $A \in \mathbb{R}^{U \times U}$. If there is $i \in U$ such that $Ae_i = \alpha e_i$ for some $\alpha \in \mathbb{R}$, then

$$\det(A) = \alpha \det(A[\{i\}^c, \{i\}^c]).$$

Proof. Let $i \in U$ be such that $Ae_i = \alpha e_i$. For every $\sigma \in \mathrm{Sym}(U)$, it holds that $A_{\sigma(i),i} = [\sigma(i) = i]\alpha$. Theorem 1.23 and the definition of determinant imply that

$$\det(A) = \det(A^\mathsf{T}) = \sum_{\sigma \in \mathrm{Sym}(U)} \mathrm{sgn}(\sigma) \prod_{k \in U} A_{\sigma(k),k} = \alpha \sum_{\sigma \in \mathrm{Sym}(U)} [\sigma(i) = i]\,\mathrm{sgn}(\sigma) \prod_{k \in U \setminus \{i\}} A_{\sigma(k),k}.$$

Let $\mathcal{S} \coloneqq \{\, \sigma \in \mathrm{Sym}(u) : \sigma(i) = i \}$. The map $(\sigma \mapsto \sigma\upharpoonright_{U \setminus \{i\}})$ from $\mathcal{S}$ into $\mathrm{Sym}(U \setminus \{i\})$ is bijective. Moreover, Proposition 1.21 ensures that for every $\sigma \in \mathcal{S}$ we have that $\mathrm{sgn}_{U \setminus \{i\}}\left(\sigma\upharpoonright_{U \setminus \{i\}}\right) = \mathrm{sgn}_U(\sigma)$. Theorem 1.23 then finishes the proof:

$$\det(A) = \alpha \left(\sum_{\tau \in \mathrm{Sym}(U \setminus \{i\})} \mathrm{sgn}(\tau) \prod_{k \in U \setminus \{i\}} A_{\tau(k),k} \right) = \alpha \det(A^\mathsf{T}[\{i\}^c, \{i\}^c]) = \alpha \det(A[\{i\}^c, \{i\}^c]). \qquad \square$$

Let U and V be finite sets, let $\phi\colon V \to U$ be a bijective function, and let $i \in V$. If we denote $j \coloneqq \phi(i)$, note that since ϕ is bijective, we have that $\phi\upharpoonright_{V \setminus \{i\}}$ is a bijection from $V \setminus \{i\}$ to $U \setminus \{j\}$. Moreover, note that for any $k \in U$, we can define $\sigma \in \mathrm{Sym}(U)$ as the transposition (jk), and have that $\sigma\phi(i) = k$. Such manipulations will become important when relating determinants with respect to ϕ with determinants of submatrices.

Proposition 2.20. Let U and V be finite sets. Let $A \in \mathbb{R}^{V \times U}$, and $i \in V$ and $j \in U$ be such that $Ae_j = \alpha e_i$, for some $\alpha \in \mathbb{R}$. Then for every bijective function $\phi\colon V \to U$ and $\sigma \in \operatorname{Sym}(U)$ such that $\sigma\phi(i) = j$, we have that

$$\det_\phi(A) = \alpha \operatorname{sgn}(\sigma) \det_\psi(A[\{i\}^c, \{j\}^c]),$$

with $\psi \coloneqq (\sigma\phi)\restriction_{V\setminus\{i\}}$.

Proof. Note that $AP_{\sigma\phi} \in \mathbb{R}^{V \times V}$ is such that $(AP_{\sigma\phi})e_i = \alpha e_i$. Hence Proposition 2.18 and Lemma 2.19 ensure that

$$\det_\phi(A) = \operatorname{sgn}(\sigma) \det_{\sigma\phi}(A) = \operatorname{sgn}(\sigma) \det(AP_{\sigma\phi}) = \operatorname{sgn}(\sigma)\alpha \det((AP_{\sigma\phi})[\{i\}^c, \{i\}^c]).$$

Moreover, note that since $Ae_j = \alpha e_i$, we have that $A[\{i\}^c, U]e_j = 0$. Proposition 1.17 ensures

$$(AP_{\sigma\phi})[\{i\}^c, \{i\}^c] = A[\{i\}^c, U]P_{\sigma\phi}[U, \{i\}^c] = A[\{i\}^c, \{j\}^c]P_{\sigma\phi}[\{j\}^c, \{i\}^c] = A[\{i\}^c, \{j\}^c]P_\psi,$$

so that

$$\det_\phi(A) = \alpha \operatorname{sgn}(\sigma) \det(AP_{\sigma\phi}[\{i\}^c, \{i\}^c]) = \alpha \operatorname{sgn}(\sigma) \det_\psi(A[\{i\}^c, \{j\}^c]). \qquad \square$$

Theorem 2.21 (Laplace Expansion)**.** Let U and V be finite sets, and let $i \in V$. Let $A \in \mathbb{R}^{V \times U}$, and let $\phi\colon V \to U$ be a bijective function. For every $j \in U$, let $\sigma_j \in \operatorname{Sym}(V)$ be such that $\sigma_j\phi^{-1}(j) = i$. Moreover, for every $j \in U$ set $\psi_j \coloneqq (\sigma_j\phi^{-1})\restriction_{U\setminus\{j\}}$. Then

$$\det_\phi(A) = \sum_{j \in U} A_{ij} \operatorname{sgn}(\sigma_j) \det_{\psi_j^{-1}}(A[\{i\}^c, \{j\}^c]).$$

Proof. Set $B \coloneqq A - e_i e_i^\mathsf{T} A$. Note that

$$A = B + \sum_{j \in U} A_{ij} e_i e_j^\mathsf{T}.$$

Theorem 1.23 implies that $\det_\phi(A) = \det(AP_\phi) = \det(P_\phi^\mathsf{T} A^\mathsf{T})$. Moreover, since $B^\mathsf{T} e_i = 0$, we have that $P_\phi^\mathsf{T} B^\mathsf{T} e_i = 0$. Furthermore,

$$\det_\phi(A) = \det(P_\phi^\mathsf{T} A^\mathsf{T}) = \det\left(P_\phi^\mathsf{T}\left(B + \sum_{j \in U} A_{ij} e_i e_j^\mathsf{T}\right)^\mathsf{T}\right) = \det\left(P_\phi^\mathsf{T} B^\mathsf{T} + \sum_{j \in U} A_{ij} P_\phi^\mathsf{T} e_j e_i^\mathsf{T}\right).$$

Finally, Proposition 1.24 and Theorem 1.23 imply

$$\det_\phi(A) = \sum_{j \in U} A_{ij} \det(P_\phi^\mathsf{T} B^\mathsf{T} + P_\phi^\mathsf{T} e_j e_i^\mathsf{T}) = \sum_{j \in U} A_{ij} \det(P_\phi^\mathsf{T}(B^\mathsf{T} + e_j e_i^\mathsf{T})) = \sum_{j \in U} A_{ij} \det((B + e_i e_j^\mathsf{T})P_\phi).$$

For every $j \in U$, note both that $(B + e_i e_j^\mathsf{T})^\mathsf{T} e_i = e_j$ and that $(B + e_i e_j^\mathsf{T})[\{i\}^c, \{j\}^c] = A[\{i\}^c, \{j\}^c]$. Hence

$$\begin{aligned}
\det_\phi(A) &= \sum_{j \in U} A_{ij} \det_\phi(B + e_i e_j^\mathsf{T}) \\
&= \sum_{j \in U} A_{ij} \det_{\phi^{-1}}((B + e_i e_j^\mathsf{T})^\mathsf{T}) \\
&= \sum_{j \in U} A_{ij} \operatorname{sgn}(\sigma_j) \det_{\psi_j}\big((B + e_i e_j^\mathsf{T})^\mathsf{T}[\{j\}^c, \{i\}^c]\big) && \text{by Proposition 2.20,} \\
&= \sum_{j \in U} A_{ij} \operatorname{sgn}(\sigma_j) \det_{\psi_j^{-1}}\big((B + e_i e_j^\mathsf{T})[\{i\}^c, \{j\}^c]\big) && \text{by Theorem 2.10,} \\
&= \sum_{j \in U} A_{ij} \operatorname{sgn}(\sigma_j) \det_{\psi_j^{-1}}(A[\{i\}^c, \{j\}^c]). && \square
\end{aligned}$$

Let V be a finite set. Let $\leq$ be a total order on V. A matrix in $\mathbb{R}^{V \times V}$ is said to be *upper triangular* (with respect to $\leq$) if $j > i$ implies that $A_{ij} = 0$. Moreover, a matrix in $\mathbb{R}^{V \times V}$ is said to be *lower triangular* (with respect to $\leq$) if its transpose is upper triangular with respect to $\leq$. A triangular matrix is any matrix that is either lower or upper triangular. Note that if A is upper triangular and $i \coloneqq \min V$, we have that

$$Ae_i = \sum_{j \in V} A_{ij} e_j = \sum_{j \in V} [j \leq i] A_{ij} e_i = A_{ii} e_i$$

This can be used, with Lemma 2.19, to give a simple formula for the determinant of triangular matrices.

Proposition 2.22. Let V be a finite set. Let $A \in \mathbb{R}^{V \times V}$ be a triangular matrix. Then

$$\det(A) = \prod_{i \in V} A_{ii}.$$

Proof. The proof is by induction on $|V|$.

If $|V| = 0$, the thesis vacuously holds. Suppose then that $|V| > 0$. Let A be a upper triangular matrix in $\mathbb{R}^{V \times V}$ with respect to $\leq$. Set $i \coloneqq \min V$. Note that $Ae_i = A_{ii} e_i$. Them Lemma 2.19 ensures

$$\det(A) = A_{ii} \det(A[\{i\}^c, \{i\}^c]).$$

Note that $A[\{i\}^c, \{i\}^c]$ is a triangular matrix with respect to $\leq$ restricted to $V \setminus \{i\}$. Therefore, the induction hypothesis ensures

$$\det(A) = A_{ii} \det(A[\{i\}^c, \{i\}^c]) = A_{ii} \prod_{j \in \{i\}^c} A_{jj} = \prod_{j \in V} A_{jj}.$$

Hence, the statement holds for upper triangular matrices. The proof is done, since if $L \in \mathbb{R}^{V \times V}$ is a lower triangular matrix with respect to $\leq$, we have that Theorem 1.23 implies that $\det(L) = \det(L^{\mathsf{T}})$, so that suffices to apply the result proved above to L^{T}. □

Lemma 2.23 (Matrix Determinant Lemma)**.** Let V be a finite set. Let $A \in \mathbb{R}^{V \times V}$ be an invertible matrix, and let $x, y \in \mathbb{R}^V$. Then

$$\det(A + xy^{\mathsf{T}}) = \det(A)(1 + y^{\mathsf{T}} A^{-1} x).$$

Proof. Let $U \coloneqq V \cup \{k\}$, for some $k \notin V$. Consider the following expression as a product of matrices in $\mathbb{R}^{U \times U}$, and written such that the first rows are indexed by V. Computation of the matrix product ensures that

$$\begin{bmatrix} I & x \\ 0 & 1 + y^{\mathsf{T}} x \end{bmatrix} = \begin{bmatrix} I & 0 \\ y^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} I + xy^{\mathsf{T}} & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ -y^{\mathsf{T}} & 1 \end{bmatrix}$$

Proposition 2.22 ensures the determinant of the LHS is $1 + y^{\mathsf{T}} x$. Proposition 2.17 ensures the determinant of the RHS is the product of the determinant of each matrix. Moreover, the first and the third matrices are triangular, so that Proposition 2.22 applies once again, and ensure both determinants are 1. For the determinant of the matrix in the middle, note that

$$\begin{bmatrix} I + xy^{\mathsf{T}} & x \\ 0 & 1 \end{bmatrix} e_k = \begin{bmatrix} I + xy^{\mathsf{T}} & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = e_k.$$

Therefore, Lemma 2.19 ensures its determinant is equal to $\det(I + xy^{\mathsf{T}})$. Therefore, the equality above implies that $\det(I + xy^{\mathsf{T}}) = 1 + y^{\mathsf{T}} x$. Proposition 2.17 then finishes the proof

$$\det\bigl(A + xy^{\mathsf{T}}\bigr) = \det(A) \det\bigl(I + A^{-1} xy^{\mathsf{T}}\bigr) = \det(A)\bigl(1 + y^{\mathsf{T}} A^{-1} x\bigr).$$ □

2.3 Matrix Tree Theorems

This section aims to prove Tutte's Matrix Tree Theorem 2.34, and use it to prove Kirchhoff's result 2.35 as a corollary. Both results arise from an interplay between linear algebra and combinatorics. Therefore, to properly understand the material, we must observe how properties from one world manifest themselves in another. For this reason, most theorems in this section are algebraic conclusions made from combinatorial hypothesis. The results build up into Proposition 2.29 and Proposition 2.32, which reverse the logic and give combinatorial conclusions from algebraic hypotesis. These results, along with Theorem 2.16, are the tools to prove Tutte's Matrix Tree Theorem.

Proposition 2.24. Let $D = (V, A, \psi)$ be a digraph, and denote by G be the underlying graph of D. If $i \in V$ and $j \in V$ are in the same component in G, then there exists $x \in \mathbb{R}^A$ such that

$$B_D x = e_j - e_i.$$

Proof. Since i and j are in the same component in G, there exists an ij-walk (in G). Let then $(i, f_1, \dots, f_\ell, j)$ be one such walk, and denote by ℓ its length. We proceed by induction on ℓ.

If $\ell = 0$, then $x \coloneqq 0$ suffices.

If $\ell > 0$, denote by k the vertex u_1 and apply the induction hypothesis to $(k, f_2, \dots, f_\ell, j)$. Therefore, there is $y \in \mathbb{R}^A$ such that $B_D y = e_j - e_k$. At least one of ik or ki is in A. In the former case, take $x \coloneqq y + e_{ik}$. In the latter case, take $x \coloneqq y - e_{ki}$. □

The proposition just proved is the connection from linear algebra and combinatorics that will be exploited into the bigger results in this section.

Let $G = (V, E)$ be a graph. A function $f \in \mathbb{R}^V$ is said to be *constant in every component of* G if, for every component C of G, there exists a value $\alpha_C \in \mathbb{R}$ such that $f\restriction_C = \alpha_C$. This is equivalent to say that $i \in V$ and $j \in V$ connected in G implies that $f(i) = f(j)$. Moreover, if $\mathcal{C} \subseteq \mathcal{P}(V)$ is the set of components of V, we have that

$$f = \sum_{C \in \mathcal{C}} \alpha_C \mathbb{1}_C.$$

Hence, f is constant in every component of C if and only if it belongs to the space spanned by the vectors $\{\, \mathbb{1}_C : C \in \mathcal{C}\}$.

Proposition 2.25. Let $D = (V, A, \psi)$ be a digraph, and let G be the underlying graph of D. Then

$$\operatorname{Null}(B_D^\mathsf{T}) = \operatorname{span}(\{\, \mathbb{1}_C : C \text{ is a component of } G.\}).$$

Proof. Suppose $B_D^\mathsf{T} f = 0$. Let $i \in V$ and $j \in V$ be vertices in the same component of G. Proposition 2.24 then ensures that there exists $x \in \mathbb{R}^A$ such that $B_D x = e_j - e_i$. Therefore

$$f(j) - f(i) = (e_j - e_i)^\mathsf{T} f = (B_D x)^\mathsf{T} f = x^\mathsf{T} B_D^\mathsf{T} f = 0.$$

Hence f is constant in every component of G. Therefore, $f \in \operatorname{span}(\{\, \mathbb{1}_C : C \text{ is a component of } G\})$.

Conversely, suppose $f \in \operatorname{span}(\{\, \mathbb{1}_C : C \text{ is a component of } G\})$. For every $ij \in A$, we have that i and j are vertices in the same component in G. Hence, $f(j) - f(i) = 0$. Therefore, for every $ij \in A$,

$$e_{ij}^\mathsf{T}(B_D^\mathsf{T} f) = (e_j - e_i)^\mathsf{T} f = f(j) - f(i) = 0.$$

Hence, $f \in \operatorname{Null}(B_D^\mathsf{T})$. □

Proposition 2.26. Let $D = (V, A, \psi)$ be a digraph with $|V| - 1$ arcs. Let G be the underlying graph of D, let $r \in V$, and let $\phi \colon V \setminus \{r\} \to A$ be a bijective function. If G is not connected, then

$$\det_\phi(B_D[\{r\}^c, A]) = 0.$$

Proof. Theorem 2.10 and Proposition 2.8 ensure

$$\det_{\phi}(B_D[\{r\}^c, A]) = \det_{\phi^{-1}}(B_D^{\mathsf{T}}[A, \{r\}^c]) = \det(B_D^{\mathsf{T}}[A, \{r\}^c]P_{\phi^{-1}}) = \det(P_{\phi^{-1}}B_D^{\mathsf{T}}[A, \{r\}^c]).$$

Proposition 1.32 reduces the statement into finding a nonzero element in $\operatorname{Null}(P_{\phi^{-1}}B_D^{\mathsf{T}}[A, \{r\}^c])$. It suffices to find a nonzero element in $\operatorname{Null}(B_D^{\mathsf{T}}[A, \{r\}^c])$. Let $C \subsetneq V$ be the component of G containing r. Proposition 2.25 ensures that $B_D^{\mathsf{T}}\mathbb{1}_{V\setminus C} = 0$.

Since $r \notin V \setminus C$, we have that $\mathbb{1}_{V\setminus C}(r) = 0$, so that Propositon 1.17 ensures

$$0 = B_D^{\mathsf{T}}\mathbb{1}_{V\setminus C} = B_D^{\mathsf{T}}[A, \{r\}^c]\mathbb{1}_{V\setminus C}.$$

Note that $\mathbb{1}_{V\setminus C}$ is nonzero because $C \subsetneq V$. Hence, the proof is done. □

Let $D = (V, A, \psi)$ be an r-arborescence with at least 2 vertices. Let $i \in V$ be a leaf in D, and denote by ji be the only arc incident on i. Note then that

$$\begin{aligned} H_D[V \setminus \{r, i\}, A - ji] &= H_{D-i}[\{r\}^c, A(D - i)] \\ T_D[V \setminus \{r, i\}, A - ji] &= T_{D-i}[\{r\}^c, A(D - i)]. \end{aligned} \tag{2.27}$$

Since the incidence, adjacency, and Laplacian matrices are defined using the head and tail matrices, this equation relates all of them in a similar manner. Also, since $D - i$ is an r-arborescence with 1 vertex fewer than D, these relations are quite useful in inductive proofs.

Proposition 2.28. Let $D = (V, A, \psi)$ be a digraph with $|V| - 1$ arcs. Let G be the underlying graph of D, let $r \in V$, and let $\phi\colon V \setminus \{r\} \to A$ be a bijective function. If G is a tree, then

$$\det_{\phi}(B_D[\{r\}^c, A])^2 = 1.$$

Proof. The proof is by induction on $|V|$.

If $|V| = 1$, the matrix $B_D[\{r\}^c, A]$ is empty, and has determinant 1. Therefore, the statement holds, since the single vertex with empty set of edges is a tree.

Let $|V| > 1$. Theorem 1.8 ensures there exists $i \in V \setminus \{r\}$ with degree 1 in G. Let $ij \in E$ be the only edge adjacent to i. Suppose $ij \in A$ — the case in which $ji \in A$ is analogous.

Note that $B_D^{\mathsf{T}}[A, \{r\}^c]e_i = e_{ij}$. Let $\sigma \in \operatorname{Sym}(V)$ be any permutation such that $\sigma\phi^{-1}(ij) = i$, and denote by $\varphi := (\sigma\phi^{-1})\restriction_{A\setminus\{ij\}}$. Then

$$\begin{aligned} \det_{\phi}(B_D[\{r\}^c, A]) &= \det_{\phi^{-1}}(B_D^{\mathsf{T}}[A, \{r\}^c]) && \text{by Theorem 2.10,} \\ &= \operatorname{sgn}(\sigma)\det_{\varphi}(B_D^{\mathsf{T}}[A - ij, V \setminus \{r, i\}]) && \text{by Proposition 2.20,} \\ &= \operatorname{sgn}(\sigma)\det_{\varphi}(B_{D-i}^{\mathsf{T}}[A(D - i), \{r\}^c]) && \text{by Equation (2.27),} \\ &= \operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(B_{D-i}[\{r\}^c, A(D - i)]) && \text{by Theorem 2.10.} \end{aligned}$$

Since i is a leaf in G, the graph $G - i$ is a tree, and $D - i$ is an orientation of it. Hence, the induction hypothesis applied to $D - i$ and φ^{-1} completes the proof:

$$\begin{aligned} \left(\det_{\phi}(B_D[\{r\}^c, A])\right)^2 &= \left(\operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(B_{D-i}[\{r\}^c, A(D - i)])\right)^2 \\ &= \operatorname{sgn}(\sigma)^2 \det_{\varphi^{-1}}(B_{D-i}[\{r\}^c, A(D - i)])^2 = 1. \end{aligned}$$ □

Proposition 2.29. Let $D = (V, A, \psi)$ be a digraph with $|V| - 1$ arcs. Let G be the underlying graph of D, let $r \in V$, and let $\phi\colon V \setminus \{r\} \to A$ be a bijective function. Then

$$\det_{\phi}(B_D[\{r\}^c, A])^2 = [G \text{ is a tree}].$$

Proof. Proposition 2.28 ensures that if G is a tree, then $\det_\phi(B_D[\{r\}^c, A])^2 = 1$. It remains to show that if G is not a tree, then $\det_\phi(B_D[\{r\}^c, A])^2 = 0$.

We proceed to prove the contrapositive of this statement. Suppose then that $\det(B_D[\{r\}^c, A])^2 \neq 0$. The contrapositive of Proposition 2.26 ensures that G is connected. Since G has $|V| - 1$ edges, then G is a tree. □

Proposition 2.29 is an algebraic criterion to determine if the underlying graph of a digraph is a tree. It bring us close to algebraically characterizing arborescences. We continue on this path with the following theorem.

Proposition 2.30. Let $D = (V, A, \psi)$ be a digraph with $|V| - 1$ arcs. Let $r \in V$, and let $\phi\colon V \setminus \{r\} \to A$ be a bijective function. Then $\det_\phi(H_D[\{r\}^c, A]) \neq 0$ implies that for every $i \in V$,

$$\left|\delta^{\text{in}}(i)\right| = [r \neq i].$$

Proof. Theorem 2.10 and Proposition 2.8 ensure

$$\det_{\phi}(H_D[\{r\}^c, A]) = \det_{\phi^{-1}}(H_D^{\mathsf{T}}[A, \{r\}^c]) = \det(H_D^{\mathsf{T}}[A, \{r\}^c]P_{\phi^{-1}}) = \det(P_{\phi^{-1}}H_D^{\mathsf{T}}[A, \{r\}^c]).$$

If $i \neq r$ is a vertex with indegree zero, then e_i is a nonzero vector in $\operatorname{Null}(H_D^{\mathsf{T}}[A, \{r\}^c])$, which ensures the determinant is zero. Therefore, if the determinant is nonzero, every vertex different from r has indegree at least 1. But since there are $|V| - 1$ arcs, every vertex different from r has indegree precisely 1. □

Proposition 2.31. Let $D = (V, A, \psi)$ be an r-arborescence. Let $\phi\colon V \setminus \{r\} \to A$ be a bijective function. Then

$$\det_{\phi}(H_D[\{r\}^c, A]) = \det_{\phi}(B_D[\{r\}^c, A]).$$

Proof. Suppose $D = (V, A, \psi)$ and ϕ are a minimal counterexample, i.e., a minimal r-arborescence such that both determinants differ. It is impossible for $|V|$ to be one, since in such case $B_D = H_D = 0$.

Hence, we have that $|V| \geq 2$. Theorem 1.9 ensures there is a vertex distinct from r with outdegree zero. Let i be one such vertex. Denote by ji the only arc incident on i. Note that $H_D^{\mathsf{T}}[A, \{r\}^c]e_i = e_{ji}$. Let $\sigma \in \operatorname{Sym}(V)$ be any permutation such that $\sigma\phi^{-1}(ji) = i$, and denote by $\varphi \coloneqq (\sigma\phi^{-1})\restriction_{A \setminus \{ji\}}$. Then

$$\begin{aligned}
\det_{\phi}(H_D[\{r\}^c, A]) &= \det_{\phi^{-1}}(H_D^{\mathsf{T}}[A, \{r\}^c]) && \text{by Theorem 2.10,}\\
&= \operatorname{sgn}(\sigma)\det_{\varphi}(H_D^{\mathsf{T}}[A - ji, V \setminus \{r, i\}]) && \text{by Proposition 2.20,}\\
&= \operatorname{sgn}(\sigma)\det_{\varphi}(H_{D-i}^{\mathsf{T}}[A(D - i), \{r\}^c]) && \text{by Equation (2.27),}\\
&= \operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(H_{D-i}[\{r\}^c, A(D - i)]) && \text{by Theorem 2.10.}
\end{aligned}$$

The same reasoning, with the same parameters σ and φ, ensures that

$$\det_{\phi}(B_D[\{r\}^c, A]) = \operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(B_{D-i}[\{r\}^c, A(D - i)]).$$

Since the statement holds for $D - i$ and φ^{-1}, we can conclude that

$$\begin{aligned}
\det_{\phi}(B_D[\{r\}^c, A]) &= \operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(B_{D-i}[\{r\}^c, A(D - i)])\\
&= \operatorname{sgn}(\sigma)\det_{\varphi^{-1}}(H_{D-1}[\{r\}^c, A(D - i)])\\
&= \det_{\phi}(H_D[\{r\}^c, A]).
\end{aligned}$$

But this contradicts the fact that D is a counterexample, and the proof is finished. □

Proposition 2.32. Let $D = (V, A, \psi)$ be a digraph with $|V| - 1$ arcs. Let $r \in V$, and let $\phi\colon V \setminus \{r\} \to A$ be a bijective function. Then

$$\det_{\phi}(H_D[\{r\}^c, A]) \det_{\phi}(B_D[\{r\}^c, A]) = [D \text{ is an } i\text{-arborescence}].$$

Proof. If D is an r-arborescence, Proposition 2.29 and Proposition 2.31 ensures that the product of determinants is one.

If D is not an r-arborescence, then either its underlying graph is not a tree, or the indegrees are not correct. If the underlying graph of D is not a tree, Proposition 2.29 ensures that $\det_{\phi}(B_D[\{r\}^c, A])^2 = 0$. This implies that $\det_{\phi}(B_D[\{r\}^c, A]) = 0$. If the degrees are not correct, then Proposition 2.30 implies that $\det(H_D[\{r\}^c, A]) = 0$. Either way, if D is not an r-arborescence, $\det_{\phi}(H_D[\{r\}^c, A]) \det_{\phi}(B_D[\{r\}^c, A]) = 0$, which concludes the proof. □

Proposition 2.33. Let V be a finite set, and suppose $S \subseteq V$. For every matrix $A \in \mathbb{R}^{V \times V}$ and $w \in \mathbb{R}^A$,

$$\det((\operatorname{Diag}(w)A)[S, S]) = \det(\operatorname{Diag}(w)[S, S]) \det(A[S, S]).$$

Proof. Set $\mathcal{T} := \binom{V}{|S|}$. For every $T \in \mathcal{T} \setminus \{S\}$, let $\phi_T\colon T \to S$ be any bijective function. Moreover, define ϕ_S to be the identity function on S. Theorem 2.16 ensures

$$\begin{aligned}\det((\operatorname{Diag}(w)A)[S, S]) &= \det(\operatorname{Diag}(w)[S, V]A[V, S])\\ &= \sum_{T \in \mathcal{T}} \det_{\phi_T^{-1}}(\operatorname{Diag}(w)[S, T]) \det_{\phi_T}(A[T, S])\\ &= \sum_{T \in \mathcal{T}} \det_{\phi_T}(\operatorname{Diag}(w)[T, S]) \det_{\phi_T}(A[T, S]).\end{aligned}$$

Hence, it is enough to show that for every $T \in \mathcal{T}$, if $T \neq S$ then $\det_{\phi_T}(\operatorname{Diag}(w)[T, S]) = 0$. Note first that $|T| = |S|$. Therefore, $S = T$ holds if and only if $S \subseteq T$.

Suppose $S \nsubseteq T$. Let $i \in T \setminus S$, and set $j := \phi_T^{-1}(i)$. Then

$$\operatorname{Diag}(w)[T, S]P_{\phi_T} e_j = \operatorname{Diag}(w)[T, S]e_i = w(i)e_i[T, 1] = 0.$$

Therefore, e_j is a nonzero vector in $\operatorname{Null}(\operatorname{Diag} w[T, S])$. Proposition 1.32 implies the determinant is zero. Hence

$$\det((\operatorname{Diag}(w)A)[S, S]) = \det_{\phi_S}(\operatorname{Diag}(w)[S, S]) \det_{\phi_S}(A[S, S]).$$

Since ϕ_S was chosen to be the identity, the proof is finished. □

Theorem 2.34 (Tutte's Matrix Tree Theorem)**.** Let $D = (V, A, \psi, w)$ be a weighted digraph. Let $r \in V$. Then

$$\det(L_D[\{r\}^c, \{r\}^c]) = \sum_{S \in \mathcal{T}_D(r)} \prod w(S).$$

Proof. For every $S \in \binom{A}{|V|-1}$, let $\phi_S\colon S \to V \setminus \{i\}$ be a bijective function. For every $S \in \binom{A}{|V|-1}$, note that

$$(\operatorname{Diag}(w)B_D^{\mathsf{T}})[S, \{r\}^c]P_{\phi_S} = (\operatorname{Diag}(w)B_D^{\mathsf{T}}P_{\phi_S})[S, S].$$

Hence, Proposition 2.33 ensures that

$$\det_{\phi_S}((\operatorname{Diag}(w)B_D^{\mathsf{T}})[S, \{r\}^c]) = \det(\operatorname{Diag}(w)[S, S]) \det(B_D^{\mathsf{T}}P_{\phi_S}[S, S]) = \det(\operatorname{Diag}(w)[S, S]) \det_{\phi_S}(B_D^{\mathsf{T}}[S, \{r\}^c]).$$

Apply this result, with both Proposition 1.16 and Theorem 2.16, to conclude that

$$\begin{aligned}
\det(L_D[\{r\}^c,\{r\}^c]) &= \det((H_D \operatorname{Diag}(w) B_D^{\mathsf{T}})[\{r\}^c,\{r\}^c]) \\
&= \sum_{S\in\binom{A}{|V|-1}} \underset{\phi_S^{-1}}{\det}(H_D[\{r\}^c,S])\, \underset{\phi_S}{\det}((\operatorname{Diag}(w)B_D^{\mathsf{T}})[S,\{r\}^c]) \\
&= \sum_{S\in\binom{A}{|V|-1}} \underset{\phi_S^{-1}}{\det}(H_D[\{r\}^c,S]) \det(\operatorname{Diag}(w)[S,S])\, \underset{\phi_S}{\det}(B_D^{\mathsf{T}}[S,\{r\}^c]) \\
&= \sum_{S\in\binom{A}{|V|-1}} \underset{\phi_S^{-1}}{\det}(H_D[\{r\}^c,S])\, \underset{\phi_S^{-1}}{\det}(B_D[\{r\}^c,S]) \det(\operatorname{Diag}(w)[S,S]) \\
&= \sum_{S\in\mathcal{T}_D(i)} \det(\operatorname{Diag}(w)[S,S]) \\
&= \sum_{S\in\mathcal{T}_D(i)} \prod_{e\in S} w(e) = \sum_{S\in\mathcal{T}_D(i)} \prod w(S).
\end{aligned}$$

Proposition 2.32 is used in the change of summation index. □

Theorem 2.35 (Kirchhoff's Matrix Tree Theorem)**.** Let $G = (V, E, \psi, w)$ be a weighted graph. Let $r \in V$. Then

$$\det(L_G[\{r\}^c,\{r\}^c]) = \sum_{S\in\mathcal{T}_G} \prod w(S).$$

Proof. Let D be the symmetric digraph of G. Proposition 1.57 ensures that $L_G = L_D$. Therefore, Theorem 2.34 and Proposition 2.1 finish the proof.

$$\det(L_G[\{r\}^c,\{r\}^c]) = \det(L_D[\{r\}^c,\{r\}^c]) = \sum_{S\in\mathcal{T}_D(r)} \prod w(S) = \sum_{S\in\mathcal{T}_G} \prod w(S).$$ □

2.4 The Algorithm

Definition 2.36. Let $D = (V, A, \psi, w)$ be a weighted digraph, and let $r \in V$. Define

$$\Phi(D,r) := \sum_{T\in\mathcal{T}_D(r)} \prod w(T).$$

Likewise, if $G = (V, A, \psi, w)$ is a weighted graph, define

$$\Phi(G) := \sum_{T\in\mathcal{T}_D} \prod w(T).$$

Proposition 2.37. Let $D = (V, A, \psi, w)$ be a weighted digraph, and let $r \in V$. Then

$$\Phi(D,r) = \det(L_D[\{r\}^c,\{r\}^c]).$$

Proof. Apply Theorem 2.34. □

The following propositions relate the problem of sampling an arborescence in a digraph with the same problem in a smaller digraph. They hint at both the recursive definition of the Naive Algorithm and its inductive proof of correctness.

Proposition 2.38. Let $D = (V, A, \psi, w)$ be a weighted digraph, let $r \in V$ and let $a_0 \in \delta^{\text{out}}(r)$ be a nonloop. Then for every $F \subseteq A \setminus \{a_0\}$,

$$\left|{\delta^{\text{in}}}_{D[F\cup\{a_0\}]}(k)\right| = [k \neq r]$$

holds for every $k \in V$ if and only if for every $k \in V(D/a_0)$

$$\left|{\delta^{\text{in}}}_{D/a_0}(k)\right| = [k \neq a_0].$$

Proof. Since a_0 is not a loop, let $i \in V$ be such that $\psi(a_0) = ri$. For every $k \in V \setminus \{r, i\}$, the set of arcs incident on k is the same on $D[F \cup \{a_0\}]$ and $D/a_0[F]$. Hence, for every $k \in V \setminus \{r, i\}$,

$$\delta^{\text{in}}{}_{D[F\cup\{a_0\}]}(k) = \delta^{\text{in}}{}_{D/a_0[F]}(k), \tag{2.39}$$

Suppose then that for every $k \in V$, it holds that $\left|\delta^{\text{in}}{}_{D[F\cup\{a_0\}]}(k)\right| = [k \neq r]$. For every $k \in V \setminus \{r, i\}$, Equation (2.39) ensures that $\left|\delta^{\text{in}}{}_{D/a_0[F]}(k) = 1\right|$. Suffices then to show that $\delta^{\text{in}}{}_{D/a_0[F]}(a_0) = \varnothing$. However, we have that

$$\delta^{\text{in}}{}_{D/a_0[F]}(a_0) = \{\, a \in F : \psi(a) \in \{jr, ji\} \text{ for some } j \in V \}.$$

Since a_0, which is not in F, is the only arc pointing to either r or i, we conclude the first half of the proof.

Supose now that for every $k \in V(D/a_0)$, it holds that $\left|\delta^{\text{in}}{}_{D/a_0[F]}(k)\right| = [k \neq a_0]$. Equation (2.39) ensures the thesis holds for every vertex in $V \setminus \{r, i\}$, and the proof is done since a_0 is incident on i, so that $\delta^{\text{in}}{}_{D[F\cup\{a_0\}]}(i) = 1$ and $\delta^{\text{in}}{}_{D[F\cup\{a_0\}]}(r) = 0$. □

Proposition 2.40. Let $D = (V, A, \psi, w)$ be a weighted digraph, let $r \in V$ and let $a_0 \in \delta^{\text{out}}(r)$ be a nonloop. For every $F \subseteq A \setminus \{a_0\}$, the underlying graph of $D[F \cup \{a_0\}]$ is connected if and only if the underlying graph of $D/a_0[F]$ is connected.

Proof. First, note that if $D[F]$ has a connected underlying graph, then so do both $D[F \cup \{a_0\}]$ and $D/a_0[F]$, and the thesis holds.

If $F \subseteq A \setminus \{a_0\}$ is such that the underlying graph of $D[F]$ is not connected, but at least one of $D[F \cup \{a_0\}]$ or $D/a_0[F]$ is connected, it holds that the underlying graph of $D[F]$ has two components, one with r and one with i. In this case, both $D[F \cup \{a_0\}]$ and $D/a_0[F]$ are connected. □

Proposition 2.41. Let $D = (V, A, \psi, w)$ be a weighted digraph, let $r \in V$ and let $a_0 \in \delta^{\text{out}}(r)$ be a nonloop. Let $\mathcal{S} \coloneqq \{\, T \in \mathcal{T}_D(r) : a_0 \in T \}$. Then $\phi \colon \mathcal{T}_{D/a_0}(a_0) \to \mathcal{S}$ defined on every $T \in \mathcal{T}_{D/a_0}(a_0)$ as

$$\phi(T) \coloneqq T \cup \{a_0\}$$

is bijective.

Proof. Apply Proposition 2.40 and Proposition 2.38 to conclude that a set $T \subseteq A \setminus \{a_0\}$ is in $\mathcal{T}_{D/a_0}(r)$ if and only if it is in $\mathcal{S}$. □

Proposition 2.42. Let $D = (V, A, \psi, w)$ be a weighted digraph, let $r \in V$, and let $a \in \delta^{\text{out}}(r)$ be a nonloop. Then

$$\Phi(D, r) = w(a)\,\Phi(D/a, a) + \Phi(D - a, r).$$

Proof. Theorem 2.35 and Proposition 2.41 ensure that

$$\begin{aligned}
\Phi(D, r) &= \sum_{T \in \mathcal{T}_D(r)} \prod w(T) \\
&= \sum_{T \in \mathcal{T}_D(r)} ([a \in T] + [a \notin T]) \prod w(T) \\
&= \sum_{T \in \mathcal{T}_D(r)} [a \in T] \prod w(T) + \sum_{T \in \mathcal{T}_D(r)} [a \notin T] \prod w(T) \\
&= w(a) \left(\sum_{T \in \mathcal{T}_D(r)} [a \in T] \prod w(T \setminus \{a\}) \right) + \left(\sum_{T \in \mathcal{T}_{D-a}(r)} \prod w(T) \right) \\
&= w(a) \left(\sum_{T \in \mathcal{T}_{D/a}(a)} w(T) \right) + \left(\sum_{T \in \mathcal{T}_{D-a}(a)} w(T) \right).
\end{aligned}$$

Note that we have also used the obvious fact that a does not belong to an r-arborescence of D if and only if it is an r-arborescence of $D - a$. □

Loops are indeed a special case. Given a weighted digraph $D = (V, A, \psi, w)$, for any vertex $r \in V$ and loop $a_0 \in \delta^{\text{out}}(r)$, the function that maps r into a_0 and fixes every other vertex is a graph isomorphism between $D - a_0$ and D/a_0. Moreover, no arborescence contains a_0. Therefore,

$$\Phi(D, r) = \Phi(D - a_0, r) = \Phi(D/a_0, a_0),$$

so Proposition 2.42 could not possibly include this case.

Loops aside, Proposition 2.42 gives the probability for an edge to belong to a random arborescence. This is the key idea in our first algorithm. However, to proper formalize the argument, we must handle a technicality and a question about the nature of randomness itself.

First things first. It is impossible for a computer, a deterministic tool, to produce randomness. The approach will be to embrace such a limitation, not to fight against it. The algorithm will actually be defined as a random variable, whose definition relies on a suitable "randomness source". This source fits the role of a `rand` function in a programming language standard library. When programming, there is little interest on what `rand` does. Likewise, in defining our algorithm, there is no *immediate* interest on the random variable that poses as "randomness source", since it is a purely measure-theoretical structure.

To define a random variable is to define a function. In order to do so, it will be convenient to have aditional information on the arcs of the input digraph. That aditional information is a total order. Note that to require an arbitrary total order on the set of arcs is by no means a limitation.

Finally, before going to the definition, a final remark is in place. For clarity, the cases should be read like a `if-else` chain. More precisely, the order of the cases is relevant, and the algorithm chooses the first option which satisfies the corresponding condition.

Definition 2.43. Let $D \coloneqq (V, A, \psi, w)$ be a weighted graph. Let $r \in V$. Let $\leq$ be a total order on A. Let $\{ X_a : a \in A\}$ be a collection of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with uniform distribution in $[0, 1]$. The *naive algorithm* is the function $\mathcal{A}(X, D, r) \colon \Omega \to \mathcal{T}_D(r) \cup \{\perp\}$ defined as

$$\mathcal{A}(X, D, r)(\omega) \coloneqq \begin{cases} \perp, & \text{if } \Phi(D, r) = 0, & \texttt{(error case)} \\ \varnothing, & \text{if } \delta^{\text{out}}(r) = \varnothing, & \texttt{(base case)} \\ \mathcal{A}(X{\restriction}_{A\setminus\{a\}}, D - a, r), & \text{if } X_a(\omega) \leq \frac{\Phi(D-a,r)}{\Phi(D,r)}, & \texttt{(drop case)} \\ \mathcal{A}(X{\restriction}_{A\setminus\{a\}}, D/a, a) \cup \{a\}, & \text{otherwise.} & \texttt{(take case)} \end{cases}$$

where $a \coloneqq \min \delta^{\text{out}}(r)$. To simplify notation, whenever X is clear from context, it will be ommited, and the algorithm will be denoted as $\mathcal{A}(D, r) \coloneqq \mathcal{A}\Big(X{\restriction}_{A(D)}, D, r\Big)$.

The appearence of $\perp$ in the definition reflects the fact that it is possible for the algorithm to receive a digraph with many arborescences and fail to output one of them. This will not actually be a problem, but for now, $\perp$ must be carried around.

Proposition 2.44. Let $D = (V, A, \psi, w)$ be a weighted graph. Let $r \in V$. Let $\leq$ be a total order on A. Let $a = \min \delta^{\text{out}}(r)$. Let $\{ X_a : a \in A\}$ be a collection of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with uniform distribution in $[0, 1]$. Suppose

(i) if $S \in \mathcal{T}_{D/a}(a)$, then $\Big\{\mathcal{A}\Big(X{\restriction}_{A\setminus\{a\}}, D/a, a\Big) = S\Big\}_\Omega \in \mathcal{F}$, and

(ii) $\mathbb{P}\Big(\mathcal{A}\Big(X{\restriction}_{A\setminus\{a\}}, D/a, a\Big) = S\Big) = (\prod w(S))/\, \Phi(D/a, a)$.

Then for every $T \in \mathcal{T}_D(r)$ such that $a \in T$,

(1) $\{\mathcal{A}(X, D, r) = T\}_\Omega \in \mathcal{F}$, and

(2) $\mathbb{P}(\mathcal{A}(X, D, r) = T) = (\prod w(T))/\, \Phi(D, r)$.

Proof. If $a \in T$, then the set $\{\mathcal{A}(D, r) = T\}_\Omega$ is a subset of $\{a \in \mathcal{A}(D, r)\}_\Omega$. But for every $\omega \in \Omega$ such that $a \in \mathcal{A}(D, r)(\omega)$, we are dealing with `(take case)`, so that

$$\{\mathcal{A}(D, r) = T\}_\Omega = \{\mathcal{A}(D, r) = T, a \in \mathcal{A}(D, r)\}_\Omega = \left\{\mathcal{A}(D/a, a) = T \setminus \{a\}, \frac{\Phi(D - a, r)}{\Phi(D, r)} < X_a\right\}_\Omega.$$

Proposition 2.41 ensures $T \setminus \{a\}$ is an a-arborescence in D/a. Therefore, the last set in the above equation is measurable, as it is the intersection of two measurable sets. The first has its measurability ensured by hypothesis (i), and the second, by the fact that $X_a \colon \Omega \to [0,1]$ is measurable. Therefore

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(D,r) = T) &= \mathbb{P}\Big(\mathcal{A}(D/a,a) = T \setminus \{a\}, \frac{\Phi(D-a,r)}{\Phi(D,r)} < X_a\Big) \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\})\mathbb{P}\Big(\frac{\Phi(D-a,r)}{\Phi(D,r)} < X_a\Big) && \text{since the } X_a\text{'s are independent,} \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\})\Big(1 - \frac{\Phi(D-a,r)}{\Phi(D,r)}\Big), \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\})w(a)\frac{\Phi(D/a,a)}{\Phi(D,r)} && \text{by Proposition 2.42,} \\
&= \Big(\frac{\prod w(T \setminus \{a\})}{\Phi(D/a,a)}\Big)w(a)\frac{\Phi(D/a,a)}{\Phi(D,r)} && \text{by hypothesis (ii),} \\
&= \frac{\prod w(T)}{\Phi(D,r)}.
\end{aligned}$$

□

Proposition 2.45. Let $D = (V, A, \psi, w)$ be a weighted graph. Let $r \in V$. Let $\leq$ be a total order on A. Let $a = \min \delta^{\text{out}}(r)$. Let $\{ X_a : a \in A\}$ be a collection of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with uniform distribution in $[0,1]$. Suppose

(i) if $S \in \mathcal{T}_{D-a}(r)$, then $\Big\{\mathcal{A}\Big(X\restriction_{A\setminus\{a\}}, D-a, r\Big) = S\Big\}_{\Omega} \in \mathcal{F}$, and

(ii) $\mathbb{P}\Big(\mathcal{A}\Big(X\restriction_{A\setminus\{a\}}, D-a, r\Big) = S\Big) = (\prod w(S))/\,\Phi(D-a,r)$.

Then for every $T \in \mathcal{T}_D(r)$ such that $a \notin T$,

(1) $\{\mathcal{A}(X,D,r) = T\}_{\Omega} \in \mathcal{F}$, and

(2) $\mathbb{P}(\mathcal{A}(X,D,r) = T) = (\prod w(T))/\,\Phi(D,r)$.

Proof. If $a \notin T$, then the set $\{\mathcal{A}(D,r) = T\}_{\Omega}$ is a subset of $\{a \notin \mathcal{A}(D,r)\}_{\Omega}$. But for every $\omega \in \Omega$ such that $a \notin \mathcal{A}(D,i)(\omega)$, we are dealing with `(drop case)`, so that

$$\{\mathcal{A}(D,r) = T\}_{\Omega} = \{\mathcal{A}(D,r) = T, a \notin \mathcal{A}(D,r)\}_{\Omega} = \Big\{\mathcal{A}(D-a,r) = T, X_a \leq \frac{\Phi(D-a,r)}{\Phi(D,r)}\Big\}_{\Omega}.$$

The last set on the above equation is the intersection of two measurable sets, the first with its measurability assured by hypothesis (i), and the second because $X_a \colon \Omega \to [0,1]$ is a measurable.Therefore,

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(D,r) = T) &= \mathbb{P}\Big(\mathcal{A}(D-a,r) = T, X_a \leq \frac{\Phi(D-a,r)}{\Phi(D,r)}\Big) \\
&= \mathbb{P}(\mathcal{A}(D-a,r) = T)\mathbb{P}\Big(X_a \leq \frac{\Phi(D-a,r)}{\Phi(D,r)}\Big) && \text{since the } X_a\text{'s are independent} \\
&= \mathbb{P}(\mathcal{A}(D-a,r) = T)\frac{\Phi(D-a,r)}{\Phi(D,r)} \\
&= \Big(\frac{\prod w(T)}{\Phi(D-a,r)}\Big)\frac{\Phi(D-a,r)}{\Phi(D,r)} && \text{by hypothesis (ii)} \\
&= \frac{\prod w(T)}{\Phi(D,r)}.
\end{aligned}$$

□

Proposition 2.46. Let $D = (V, A, \psi, w)$ be a weighted graph, with $w \in \mathbb{R}^A_{++}$. Let $r \in V$. Let $\leq$ be a total order on A. Let $a = \min \delta^{\text{out}}(r)$. Let $\{ X_a : a \in A\}$ be a collection of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with uniform distribution in $[0,1]$. Then for every $T \in \mathcal{T}_D(r)$

(i) $\{\mathcal{A}(X,D,r)=T\}_{\Omega} \in \mathcal{F}$, and

(ii) $\mathbb{P}(\mathcal{A}(X,D,r)=T) = w(T)/\,\Phi(D,r)$.

Proof. We proceed by induction on $|A|$.

If $A = \varnothing$, since $\mathcal{T}_D(r)$ is nonempty, it follows that $\mathcal{T}_D(r) = \{\varnothing\}$. In such a case,

$$\Omega = \{\mathcal{A}(D,i) = \varnothing\}_{\Omega},$$

since the naive algorithm will output $\varnothing$ regardless of the input $\omega \in \Omega$. Since $\Omega \in \mathcal{F}$, we have that $\{\mathcal{A}(D,r)=\varnothing\}_{\Omega}$ is measurable. Moreover, since it is a probability space, its measure must be 1, which is equal to the RHS of (i), since $\Phi(D,r)=1$ and the empty product is 1.

Let then A be nonempty, and let $T \in \mathcal{T}_D(r)$ be any r-arborescence. Since $T \in \mathcal{T}_D(r)$ and $w > 0$, it follows that

$$\Phi(D,r) = \sum_{S \in \mathcal{T}_D(r)} \prod w(S) \geq w(T) > 0.$$

Moreover, $\delta^{\text{out}}(r)$ is nonempty. In such a case, let a be the minimum element of $\delta^{\text{out}}(r)$.

Observe that we can now assume that the algorithm is not on `(error case)`, since $\Phi(D,r) > 0$, nor on `(base case)`, since $\delta^{\text{out}}(r) \neq \varnothing$.

There are two cases to consider, depending on whether a is in T or not. Note that the respective hypothesis in Proposition 2.44 or Proposition 2.45 are assured by the induction hypothesis. Hence, either Proposition 2.44 or Proposition 2.45 ensures (i) and (ii) holds, depending on whether $a \in T$ or $a \notin T$, respectively. □

Theorem 2.47. Let $D = (V, A, \psi, w)$ be a weighted graph, with $w \in \mathbb{R}^{A}_{++}$. Let $r \in V$. Let $\leq$ be a total order on A. Let $a = \min\, \delta^{\text{out}}(r)$. Let $\{\, X_a : a \in A\}$ be a collection of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with uniform distribution in $[0,1]$. Then

(i) The function $\mathcal{A}(D,r)$ is a random variable;

(ii) For every $T \in \mathcal{T}_D(r)$, it holds that

$$\mathbb{P}(\mathcal{A}(D,r)=T) = \frac{\prod w(T)}{\Phi(D,r)};$$

(iii) If D has at least one r-arborescence, then $\mathbb{P}(\mathcal{A}(D,r)=\perp) = 0$.

Proof. First, note that if D has no r-arborescences, then $\Phi(D,r)=0$. Therefore,

$$\Omega = \{\mathcal{A}(D,r) = \perp\}_{\Omega},$$

since the algorithm will always go through `(error case)`, regardless of the input $\omega \in \Omega$. Since $\Omega \in \mathcal{F}$, it follows that $\{\mathcal{A}(D,i) = \perp\}_{\Omega}$ is measurable. In such a case, (i), (ii), and (iii) hold.

If D has at least one r-arborescence, Proposition 2.46 applies and ensures (ii). It remains to show that $\mathcal{A}(D,r)$ is indeed a random variable, and that (iii) holds.

To show that $\mathcal{A}(D,r)$ is a random variable, it remains only to show that the preimage of $\perp$ is measurable. Note then that

$$\{\mathcal{A}(D,r)=\perp\}_{\Omega} = \Omega \setminus \left(\bigcup_{T \in \mathcal{T}_D(r)} \{\mathcal{A}(D,r)=T\}_{\Omega} \right),$$

and the RHS is the complement of a finite union of measurable sets, so that it is indeed measurable.

Finally, using Proposition 2.37, we have that

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(D,r) = \bot) &= 1 - \sum_{T \in \mathcal{T}_D(r)} \mathbb{P}(\mathcal{A}(D,r) = T) \\
&= 1 - \left(\sum_{T \in \mathcal{T}_D(r)} \frac{\prod w(T)}{\Phi(D,r)} \right) \\
&= 1 - \frac{1}{\Phi(D,r)} \left(\sum_{T \in \mathcal{T}_D(r)} \prod w(T) \right) \\
&= 1 - \left(\frac{\Phi(D,r)}{\Phi(D,r)} \right) = 0,
\end{aligned}$$

which demonstrates (iii) and finishes the proof. □

If we let $\bot$ represent a division by zero error, the pseudocode for the algorithm writes itself. Aside from being a recusive method, it is still remarkably similar to algorithm A described by Kulkarni in [18].

```
function SAMPLE($D, r$)
    Let $D = (V, A, \psi, w)$.
    if $\delta^{\text{out}}(r) = \varnothing$ then return $\varnothing$.
    Let $a \in \delta^{\text{out}}(r)$, and let $p \leftarrow \Phi(D - a, r) / \Phi(D, r)$.
    Let $x$ be a uniform random variable in the interval $[0, 1]$.
    if $x \leq p$ then return SAMPLE($D - a, r$).
    else   return $\{a\} \cup$ SAMPLE($D/a, a$).
```

2.5 Effective Resistances as Marginal Probabilities

We now use the theory developed so far to give an alternate formula for the marginal probabilities of an edge belonging to the output of the naive algorithm.

Proposition 2.48. Let $G = (V, E, \psi, w)$ be a connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let $e_0 \in E$ be a nonloop. Denote by i and j the elements of $\psi(e_0)$. Then

$$(e_i - e_j)^\mathsf{T} L_G^\dagger (e_i - e_j) = (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}.$$

Proof. First, note that by indexing the first rows and the first columns by $V \setminus \{i\}$, the equality $L_G \mathbb{1} = 0$ turns into

$$\begin{bmatrix} L_G[\{i\}^c, \{i\}^c] & L_G[\{i\}^c, \{i\}^c] e_i \\ -e_i^\mathsf{T} L_G[\{i\}^c, \{i\}^c] & (L_G)_{ii} \end{bmatrix} \begin{bmatrix} \mathbb{1} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Solving for both $L_G e_i$ and $(L_G)_{ii}$, we conclude

$$\begin{aligned}
L_G[\{i\}^c, \{i\}^c] e_i &= -L_G[\{i\}^c, \{i\}^c] \mathbb{1} \\
(L_G)_{ii} &= \mathbb{1}^\mathsf{T} L_G[\{i\}^c, \{i\}^c] \mathbb{1}.
\end{aligned}$$

Hence,

$$L_G = \begin{bmatrix} L_G[\{i\}^c, \{i\}^c] & -L_G[\{i\}^c, \{i\}^c] \mathbb{1} \\ -\mathbb{1}^\mathsf{T} L_G[\{i\}^c, \{i\}^c] & \mathbb{1}^\mathsf{T} L_G[\{i\}^c, \{i\}^c] \mathbb{1} \end{bmatrix} = \begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix} L_G[\{i\}^c, \{i\}^c] \begin{bmatrix} I & -\mathbb{1} \end{bmatrix}.$$

Since G is connected, it has at least a spanning tree. Moreover, since w is positive, we conclude that $\det(L_G[\{i\}^c, \{i\}^c])$ is nonzero, and, therefore, $L_G[\{i\}^c, \{i\}^c]$ is invertible. Hence, it is injective and surjective. Moreover, note that

$$\begin{bmatrix} I \\ -\mathbb{1} \end{bmatrix} x = 0 \implies x = 0.$$

Therefore, this matrix is injective, and its transpose is surjective. Proposition 1.50 applied twice implies

$$L_G^\dagger = \begin{bmatrix} I & -\mathbb{1} \end{bmatrix}^\dagger L_G[\{i\}^c, \{i\}^c]^{-1} \begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}^\dagger .$$

From Example 1.52, we have that

$$\begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}^\dagger (e_i - e_j) = \begin{bmatrix} I - \frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T} & -\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T} \end{bmatrix} \begin{bmatrix} -e_j \\ 1 \end{bmatrix} = -e_j.$$

Therefore,

$$\begin{aligned} (e_i - e_j) L_G^\dagger (e_i - e_j) &= \left(\begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}^\dagger (e_i - e_j) \right)^\mathsf{T} L_G[\{i\}^c, \{i\}^c]^{-1} \left(\begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}^\dagger (e_i - e_j) \right) \\ &= (-e_j)^\mathsf{T} (L_G[\{i\}^c, \{i\}^c]^{-1})(-e_j) = (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}. \end{aligned}$$

□

Proposition 2.49. Let $G = (V, E, \psi, w)$ be a weighted, connected graph, with $w \in \mathbb{R}^E_{++}$. Let $e_0 \in E$ be a nonloop. Denote by i and j the elements of $\psi(e_0)$. Then

$$\frac{\Phi(G/e_0)}{\Phi(G)} = (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}.$$

Proof. Proposition 1.57 implies that

$$L_G = L_{G-e_0} + w(e_0)(e_i - e_j)(e_i - e_j)^\mathsf{T}.$$

Therefore,

$$L_{G-e_0}[\{i\}^c, \{i\}^c] = L_G[\{i\}^c, \{i\}^c] - w(e_0) e_j e_j^\mathsf{T}.$$

Since G is connected and $w \in \mathbb{R}^V_{++}$, we have that $L_G[\{i\}^c, \{i\}^c]$ is invertible. Hence, Lemma 2.23 ensures that

$$\det(L_{G-e_0}[\{i\}^c, \{i\}^c]) = \det(L_G[\{i\}^c, \{i\}^c]) \big(1 - w(e_0) e_j^\mathsf{T} L_G[\{i\}^c, \{i\}^c]^{-1} e_j\big).$$

Since $w(e_0) > 0$,

$$\begin{aligned} (L_G[\{i\}^c, \{i\}^c])^{-1}_{jj} &= \frac{1}{w(e_0)} \left(1 - \frac{\Phi(G - e_0)}{\Phi(G)} \right) \\ &= \frac{\Phi(G) - \Phi(G - e_0)}{w(e_0)\, \Phi(G)} \\ &= \frac{w(e_0)\, \Phi(G/e_0) + \Phi(G - e_0) - \Phi(G - e_0)}{w(e_0)\, \Phi(G)} && \text{by Proposition 2.42} \\ &= \frac{\Phi(G/e_0)}{\Phi(G)}. \end{aligned}$$

□

Both propositions just proved describe a formula to calculate the probability of an edge to belong to the output of a sampling algorithm using the pseudoinverse of the Laplacian instead of its determinant. This is interesting because it only demands 4 entries of the pseudoinverse matrix to be known.

Theorem 2.50. Let $G = (V, E, \psi, w)$ be a weighted connected graph. Let $e_0 \in E$ be such that $|\psi(e_0)| = 2$. Denote by i and j the elements of $\psi(e_0)$. Let $\mathcal{A}(G) \colon \Omega \to \mathcal{T}_G$ be a random variable such that for every $T \in \mathcal{T}_G$,

$$\mathbb{P}(\mathcal{A}(G) = T) = \frac{\prod c(T)}{\Phi(G)}.$$

Then

$$\mathbb{P}(e_0 \in \mathcal{A}(G)) = w(e_0)(e_i - e_j)^\mathsf{T} L_G^\dagger (e_i - e_j).$$

Proof. Apply both propositions just proved:

$$\mathbb{P}(e_0 \in \mathcal{A}(G)) = w(e_0) \frac{\Phi(G/e_0)}{\Phi(G)} = w(e_0)(L_G[\{i\}^c, \{i\}^c]^{-1})_{jj} = w(e_0)(e_i - e_j)^\mathsf{T} L_G^\dagger (e_i - e_j).$$

□

Chapter 3

Markov Chains

3.1 Markov Chains and Random Walks

A *stochastic process* in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function X on a totally ordered set T such that X_k is a random variable on Ω for every $k \in T$, all of which take on values in the same measurable space $(V, \mathcal{M})$. Call V the *state space* of the stochastic process. A stochastic process X with $T = \mathbb{N}$ is said to be *discrete time*. A discrete time stochastic process X is a *Markov chain* if

$$\mathbb{P}(X_{k+1} = s_{k+1} \mid X_k = s_k, \dots, X_0 = s_0) = \mathbb{P}(X_{k+1} = s_{k+1} \mid X_k = s_k) \tag{3.1}$$

for every $k \in \mathbb{N}$ and $s \colon ([k+2]-1) \to V$ such that both conditional probabilities are defined, i.e., such that $\mathbb{P}(X_k = s_k, \dots, X_0 = s_0) > 0$. Condition (3.1) is called the *Markov property.*

As mentioned in Section 1.9, if V is at most countable, the measurable space we are interested in is $(V, \mathcal{P}(V))$. Hence, whenever we state that X is a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V, we are saying that for every $t \in \mathbb{N}$ we have that $X_t \colon \Omega \to V$ is a measurable function with respect to $\mathcal{F}$ and $\mathcal{P}(V)$.

Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. The *trajectory of X* is the function $\mathrm{Traj}_X \colon \Omega \to V^{\mathbb{N}}$ defined as, for every $\omega \in \Omega$ and $t \in \mathbb{N}$,

$$(\mathrm{Traj}_X(\omega))_t \coloneqq X_t(\omega).$$

First, note that for every $n \in \mathbb{N}$ and $s \in V^{[n]-1}$,

$$\left\{\mathrm{Traj}_X{\upharpoonright}_{[n]-1} = s\right\}_\Omega = \{\,\omega \in \Omega : \mathrm{Traj}_X(\omega){\upharpoonright}_{[n]-1} = s\} = \bigcap_{t \in [n]-1} \{X_t = s_t\}_\Omega. \tag{3.2}$$

For now, Traj_X is mostly a concise way of describing a sequence of states being observed in a Markov chain. Section 3.3 will show it is much more important than that. By now, suffices to note that the definitions of stochastic process and σ-algebra ensure that every event as described above is measurable.

Let X be a Markov chain, and suppose its state space V is finite. For each $k \in \mathbb{N}$, the *transition matrix of X at time k* is the matrix $P_k \colon V \times V \to \mathbb{R}$ defined by $(P_k)_{ij} \coloneqq \mathbb{P}(X_{k+1} = j \mid X_k = i)$ for each $i, j \in V$. If $P_k = P_\ell$ for every $k, \ell \in \mathbb{N}$, the Markov chain is *time-homogeneous*, and the common value $P \colon V \times V \to \mathbb{R}$ is the *transition matrix.* In this work, every Markov chain is assumed time-homogeneous, so that it suffices to define a single $P \in \mathbb{R}^{V \times V}$ such that $P_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i)$, for every $i, j \in V$.

Transition matrices have many interesting properties. For such reason, matrices that could be transition matrices of some Markov chain are given a special name. Let V be a finite set. A matrix $P \in \mathbb{R}^{V \times V}$ is *stochastic* if

1. $P_{ij} \geq 0$ for every i and j in V,

2. $e_i^{\mathsf{T}} P \mathbb{1} = 1$ for every $i \in V$.

Theorem 3.15 will actually prove that for every stochastic matrix P there is a Markov chain whose transition matrix is P.

Let X be a time-homogeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. Let $P \in \mathbb{R}^{V \times V}$ be its transiton matrix, and define $x \in \mathbb{R}^V$ as $x_i \coloneqq \mathbb{P}(X_0 = i)$ for every $i \in V$. Note that, for every $i \in V$,

$$\begin{aligned} e_i^{\mathsf{T}} P^{\mathsf{T}} x &= \sum_{j \in V} P_{ji} x_j \\ &= \sum_{j \in V} \mathbb{P}(X_1 = i \mid X_0 = j) \mathbb{P}(X_0 = j) \\ &= \sum_{j \in V} \mathbb{P}(X_1 = i, X_0 = j) \\ &= \mathbb{P}(X_1 = i). \end{aligned}$$

Hence, the ith coordinate of $P^{\mathsf{T}} x$ is the probability that X_1 is i. An inductive argument generalizes this observation to the fact that for every $t \in \mathbb{N}$ and $i \in V$,

$$e_i^{\mathsf{T}} \left(P^{\mathsf{T}}\right)^t x = \mathbb{P}(X_t = i). \tag{3.3}$$

Let X be a time-homogeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. If X is such that for every $t \in \mathbb{N}$ and $i \in V$ it holds that $\mathbb{P}(X_t = i) = \mathbb{P}(X_0 = i)$, then X is said to be *stationary*. Let $P \in \mathbb{R}^{V \times V}$ be the transiton matrix of X. Equation (3.3) ensures that X is stationary if and only if the vector $\pi \in \mathbb{R}^V$, defined for every $i \in V$ by $\pi_i \coloneqq \mathbb{P}(X_0 = i)$, satisfies

$$P^{\mathsf{T}} \pi = \pi.$$

If this is the case, the vector π is also referred to as a *stationary distribution* of X.

Let $D = (V, A, w)$ be a weighted simple digraph, with $w \in \mathbb{R}^A_{++}$. A *random walk on* D is a Markov chain X, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V, such that for every $t \in \mathbb{N}$ and $i, j \in V$,

$$\mathbb{P}(X_{t+1} = j \mid X_t = i) = [ij \in A] \frac{w(ij)}{\sum w(\delta^{\text{out}}(i))}. \tag{3.4}$$

Note that a random walk on a graph is always a time-homogeneous Markov chain. If $G = (V, E, w)$ is a weighted simple graph, with $w \in \mathbb{R}^A_{++}$, a *random walk on* G is a random walk on its symmetric digraph. If G is not only simple and weighted, but also connected, the matrices defined on Section 1.6 make it possible to write Equation (3.4) as a matrix equation. If $G = (V, E, w)$ is simple, weighted, and connected, then to state that X is a random walk on G is to state that X is a time-homogeneous Markov chain, with finite state space V, and such that its transition matrix $P \in \mathbb{R}^{V \times V}$ satisfies

$$P = D_G^{-1} A_G.$$

Actually, it is possible to drop the hypothesis that G is connected and state the above equality as $P = D_G^{\dagger} A_G$. We do not do so, however, because it would demand more proofs to be made, and it would not improve the exposition. However, the remarks above justify the following definition.

Definition 3.5. Let G be a simple, weighted and connected graph. The *transition matrix of* G is the matrix $P \in \mathbb{R}^{V \times V}$ given by

$$P = D_G^{-1} A_G.$$

A vector $\pi \in \mathbb{R}^V_+$ is a *stationary distribution of* G if $\mathbb{1}^{\mathsf{T}} \pi = 1$ and $P^{\mathsf{T}} \pi = \pi$.

Observe that Equation (3.3) ensures that a stationary distribution of a graph G is a stationary distribution of any random walk on G.

Let X be a time-homogeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. Its *transition digraph* is the digraph $D \coloneqq (V, A, w)$, where

$$\begin{aligned} A &\coloneqq \{\, ij \in V \times V : \mathbb{P}(X_1 = j \mid X_0 = i) > 0 \}, \\ w(ij) &\coloneqq \mathbb{P}(X_1 = j \mid X_0 = i) \qquad \forall ij \in A. \end{aligned}$$

It is no surprise that a Markov chain is a random walk on its transition digraph. The interesting consequence of the above definition is that every time-homogeneous Markov chain with a finite state space can be seen as a random walk on a digraph.

Definition 3.6. A time-homogeneous Markov chain with finite state space is *irreducible* if its transition digraph is strongly connected.

Definition 3.7. Let X be a time-homogeneous Markov chain with a finite state space V. Let $P \in \mathbb{R}^{V \times V}$ be its transition matrix. The Markov chain X is *time-reversible* if there exists $\pi \in \mathbb{R}^V_{++}$ such that for every $i, j \in V$,

$$\pi_i P_{ij} = \pi_j P_{ji}.$$

Proposition 3.8. Let X be a time-homogeneous and time-reversible Markov chain with finite state space. There exists a simple graph G such that X is a random walk on G.

Proof. Let V be the state space of X. Since X is time-reversible, let $\pi \in \mathbb{R}^V_{++}$ such that for every $i, j \in V$ we have that $\pi_i P_{ij} = \pi_j P_{ji}$. Note that $P_{ij} > 0$ implies that $P_{ji} > 0$, for every i and j in V.

Define $G \coloneqq (V, E, w)$, where

$$E \coloneqq \left\{ ij \in \binom{V}{2} : P_{ij} > 0 \right\},$$

and $w \colon E \to \mathbb{R}$ is given by $w(ij) \coloneqq \pi_i P_{ij}$ for every $ij \in E$. Note that time-reversibility is used to ensure that w is well defined. Moreover,

$$\pi_i = \pi_i \cdot 1 = \pi_i \left(\sum_{j \in V} P_{ij} \right) = \sum_{j \in V} \pi_i P_{ij} = \sum_{j \in V} w(ij) = \sum w(\delta(i)),$$

Hence, for every $i, j \in V$,

$$P_{ij} = [ij \in E] \frac{w(ij)}{\pi_i} = [ij \in E] \frac{w(ij)}{\sum w(\delta(i))},$$

so that X is a random walk on G. □

Proposition 3.9. Let $G = (V, E, w)$ be a connected, weighted, and simple graph. Then

$$\pi \coloneqq \frac{1}{2w^\mathsf{T} \mathbb{1}} D_G \mathbb{1}$$

is the only stationary distribution of G.

Proof. Let $D = (V, A, w)$ be any orientation of G. Proposition 1.56 and Proposition 1.54 imply

$$\begin{aligned} \mathbb{1}^\mathsf{T} D_G \mathbb{1} &= \mathbb{1}^\mathsf{T} \big(H_D \operatorname{Diag}(w) H_D^\mathsf{T} + T_D \operatorname{Diag}(w) T_D^\mathsf{T} \big) \mathbb{1} \\ &= \mathbb{1}^\mathsf{T} H_D \operatorname{Diag}(w) H_D^\mathsf{T} \mathbb{1} + \mathbb{1}^\mathsf{T} T_D \operatorname{Diag}(w) T_D^\mathsf{T} \mathbb{1} \\ &= 2\, \mathbb{1}^\mathsf{T} \operatorname{Diag}(w) \mathbb{1} \\ &= 2w^\mathsf{T} \mathbb{1}. \end{aligned}$$

Therefore, $\mathbb{1}^\mathsf{T} \pi = 1$. Moreover, for every $i \in V$, Proposition 1.55 implies $e_i^\mathsf{T} D_G \mathbb{1} = \sum w(\delta(i))$. Since G is connected, $\delta(i) \neq \varnothing$. Furthermore, since $w \in \mathbb{R}^E_{++}$, we have that $w(\delta(i)) > 0$. Hence, for every $i \in V$, we have that $e_i^\mathsf{T} \pi > 0$.

Let $P \in \mathbb{R}^{V \times V}$ be the transition matrix of G. To conclude that $\pi \in \mathbb{R}^V$ is a stationary distribution of G, it remains only to show that $P^\mathsf{T}\pi = \pi$. Proposition 1.56 and Proposition 1.55 imply

$$P^\mathsf{T}\pi = \frac{1}{2w^\mathsf{T}\mathbb{1}}\left(D_G^{-1}A_G\right)^\mathsf{T} D_G\mathbb{1} = \frac{1}{2w^\mathsf{T}\mathbb{1}} A_G D_G^{-1} D_G \mathbb{1} = \frac{1}{2w^\mathsf{T}\mathbb{1}} A_G\mathbb{1} = \frac{1}{2w^\mathsf{T}\mathbb{1}} D_G\mathbb{1} = \pi.$$

Therefore, π as defined on the statement is a stationary distribution. Suppose now that $x \in \mathbb{R}^V$ is a stationary distribution of G. By definition, $P^\mathsf{T}x = x$, or, equivalently, $(I - P)^\mathsf{T}x = 0$. Hence Proposition 1.57 ensures that

$$\begin{aligned}
0 &= (I-P)^\mathsf{T}x \\
&= (I-P)^\mathsf{T}(D_G D_G^{-1})x \\
&= (D_G(I-P))^\mathsf{T}\left(D_G^{-1}x\right) \\
&= (D_G(I - D_G^{-1}A_G))^\mathsf{T}\left(D_G^{-1}x\right) \\
&= (D_G - A_G)^\mathsf{T}\left(D_G^{-1}x\right) \\
&= L_G(D_G^{-1}x).
\end{aligned}$$

Proposition 1.58 and the fact that G is connected then ensure that $D_G^{-1}x \in \operatorname{span}(\mathbb{1})$. In other words, there exists $\alpha \in \mathbb{R}$ such that $x = \alpha D_G\mathbb{1}$. Moreover, $\mathbb{1}^\mathsf{T}x = 1$, so that

$$1 = \mathbb{1}^\mathsf{T}(\alpha D_G \mathbb{1}) = \alpha\mathbb{1}^\mathsf{T} D_G\mathbb{1} = \alpha 2w^\mathsf{T}\mathbb{1}.$$

Hence $\alpha = (2w^\mathsf{T}\mathbb{1})^{-1}$, so that $x = \pi$. □

Corollary 3.10. Let X be time-homogeneous, irreducible, time-reversible Markov chain, with finite state space. Then X has a unique stationary distribution.

Proof. Proposition 3.8 ensures there exists G such that X is a random walk on G. Since X is irreducible, G is connected. Morover, since X is a random walk on G, Proposition 3.9 ensures that this distribution is unique. □

Note how the statement of Corollary 3.10 is purely about Markov chains, with no random walks on graphs involved.

3.2 Markov Chain Construction

We now focus on the construction of Markov chains. As in the case with the naive algorithm, it is only possible to adapt a “randomness source” into a more useful format. Therefore, once again, composition will serve as the main tool, by describing how to alter the randomness into a Markov chain. A similar construction is made in [10].

We have, however, to start somewhere. For this reason, we assume the existence of a discrete time stochastic process X, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, given by independent, uniformly distributed random variables taking values in $[0,1]$. Whenever this stochastic process exists, the space $(\Omega, \mathcal{F}, \mathbb{P})$ is *chainable*. Furthermore, let V be a finite set, let $\mu\colon [0,1] \to V$ be a measurable function, and let T be a function on V such that $T_i\colon [0,1] \to V$ is measurable for every $i \in V$. Then (X, V, T, μ) is a *chain specification in* $(\Omega, \mathcal{F}, \mathbb{P})$. As the name suggests, (X, V, T, μ) can be used to create a time-homogeneous Markov chain, since it encodes the randomness source, the state space, the transitions and the initial probability. Before going into that, though, this small result will prove itself laborsaving for the task at hand.

Proposition 3.11. Let V be a finite set. Let $x \in \mathbb{R}_+^V$ be such that $\mathbb{1}^\mathsf{T}x = 1$. Then there exists a measurable function $\nu\colon [0,1] \to V$ such that

1. for every $i \in V$, the preimage $\nu^{-1}(i)$ is an interval,
2. for every $i \in V$, it holds that $|\nu^{-1}(i)| = x_i$.

Proof. Let $\leq$ be a total order on V. Define

$$\nu(0) \coloneqq \inf V,$$

$$\left(\nu\restriction_{(0,1]}\right)^{-1}(i) \coloneqq \left(\sum_{j<i} x_j, \sum_{j\leq i} x_j\right]. \qquad \square$$

Definition 3.12. Let (X, V, T, μ) be a chain specification in $(\Omega, \mathcal{F}, \mathbb{P})$. The *Markov chain M defined by* (X, V, T, μ) is the function whose domain is $\mathbb{N}$, defined as

$$M_t(\omega) \coloneqq \begin{cases} \mu X_0(\omega), & \text{if } t = 0, \\ T_{M_{t-1}(\omega)} X_t(\omega), & \text{otherwise,} \end{cases}$$

for every $t \in \mathbb{N}$ and $\omega \in \Omega$.

Proposition 3.13. Let (X, V, T, μ) be a chain specification in $(\Omega, \mathcal{F}, \mathbb{P})$. Let M be the Markov chain defined by it, and let $t \in \mathbb{N}$. If $\mathbb{P}(M_t = i) > 0$, then

$$\mathbb{P}(M_{t+1} = j \mid M_t = i) = \mathbb{P}(T_i X_{t+1} = j).$$

Proof. Let $\mathcal{S}_t$ be $\{\mu\}$ if $t = 0$, and $\{\, T_i : i \in V\}$ otherwise. For every $t \in \mathbb{N}$,

$$\{M_t = i\}_\Omega = \bigcup_{\nu\in\mathcal{S}_t} \{\nu X_t = i\}_\Omega.$$

Let then $t \in \mathbb{N}$. Then

$$\begin{aligned}
\{M_{t+1} = j, M_t = i\}_\Omega &= \{T_i X_{t+1} = j, M_t = i\}_\Omega \\
&= \{T_i X_{t+1} = j\}_\Omega \cap \{M_t = i\}_\Omega \\
&= \{T_i X_{t+1} = j\}_\Omega \bigcap \left(\bigcup_{\nu\in\mathcal{S}_t} \{\nu X_t = i\}_\Omega\right) \\
&= \bigcup_{\nu\in\mathcal{S}_t} \{T_i X_{t+1} = j\}_\Omega \cap \{\nu X_t = i\}_\Omega \\
&= \bigcup_{\nu\in\mathcal{S}_t} \left\{X_{t+1} \in T_i^{-1}(j)\right\}_\Omega \cap \left\{X_t \in \nu^{-1}(i)\right\}_\Omega
\end{aligned}$$

Since $\{\, X_t : t \in \mathbb{N}\}$ is an independent family of functions, and $\mathcal{S}_t$ is finite,

$$\begin{aligned}
\mathbb{P}(M_{t+1} = j, M_t = i) &= \sum_{\nu\in\mathcal{S}_t} \mathbb{P}\big(X_{t+1} \in T_i^{-1}(j)\big)\mathbb{P}\big(X_t \in \nu^{-1}(i)\big) \\
&= \mathbb{P}(T_i X_{t+1} = j)\left(\sum_{\nu\in\mathcal{S}_t} \mathbb{P}(\nu X_t = i)\right) \\
&= \mathbb{P}(T_i X_{t+1} = j)\mathbb{P}\left(\bigcup_{\nu\in\mathcal{S}_t} \{\nu X_t = i\}_\Omega\right) \\
&= \mathbb{P}(T_i X_{t+1} = j)\mathbb{P}(M_t = i).
\end{aligned}$$

Therefore, if $\mathbb{P}(M_t = i) > 0$,

$$\mathbb{P}(M_{t+1} = j \mid M_t = i) = \mathbb{P}(T_i X_{t+1} = j). \qquad \square$$

Up until this point, the expression *Markov chain defined by the chain specification* (X, V, T, μ) is merely foreshadowing. We proceed in showing it is not a misnomer.

Proposition 3.14. Let (X, V, T, μ) be a chain specification in $(\Omega, \mathcal{F}, \mathbb{P})$. Let M be the Markov chain defined by it. Then M is indeed a Markov chain. Moreover, it is time-homogeneous and has a finite state space.

Proof. For every $t \in \mathbb{N}$, the function M_t is a composition of measurable functions, and, as such, is itself measurable. Remains to prove the Markov property (3.1). Note that for every $k \in \mathbb{N}$ and for every $x \in V^{[k+2]-1}$

$$\begin{aligned}\Big\{\mathrm{Traj}_M\restriction_{[k+2]-1} = x\Big\}_\Omega &= \bigcap_{t \in [k+2]-1} \{M_t = x_t\}_\Omega \\ &= \{M_0 = x_0\}_\Omega \bigcap \left(\bigcap_{t=1}^{k+1} \{M_t = x_t\}_\Omega\right) \\ &= \{\mu X_0 = x_0\}_\Omega \bigcap \left(\bigcap_{t=1}^{k+1} \{T_{x_{t-1}} X_t = x_t\}_\Omega\right) \\ &= \left\{X_0 \in \mu^{-1}(x_0)\right\}_\Omega \bigcap \left(\bigcap_{t=1}^{k+1} \Big\{X_t \in T_{x_{t-1}}^{-1}(x_t)\Big\}_\Omega\right).\end{aligned}$$

Proposition 3.13 and the fact that $\{X_t\}$ is an independent family of random variables then imply

$$\begin{aligned}\mathbb{P}\Big(\mathrm{Traj}_M\restriction_{[k+2]-1} = x\Big) &= \mathbb{P}\big(X_0 \in \mu^{-1}(x_0)\big) \prod_{t=1}^{k+1} \mathbb{P}\Big(X_t \in T_{x_{t-1}}^{-1}(x_t)\Big) \\ &= \mathbb{P}(M_0 = x_0) \prod_{t=1}^{k+1} \mathbb{P}(T_{x_{t-1}} X_t = x_t) \\ &= \mathbb{P}(M_0 = x_0) \prod_{t=1}^{k+1} \mathbb{P}(M_t = x_t \mid M_{t-1} = x_{t-1}),\end{aligned}$$

as long as $\mathbb{P}(\mathrm{Traj}_M\restriction_{[k+1]-1} = x\restriction_{[k+1]-1}) > 0$ — so that all the conditional probabilities are well defined.

To finish the proof, suffices to apply this equality twice. Let $k \in \mathbb{N}$. Then

$$\begin{aligned}\mathbb{P}\Big(\mathrm{Traj}_M\restriction_{[k+2]-1} = x\Big) &= \mathbb{P}(M_0 = x_0) \prod_{t=1}^{k+1} \mathbb{P}(M_t = x_t \mid M_{t-1} = x_{t-1}) \\ &= \mathbb{P}(M_{k+1} = x_{k+1} \mid M_k = x_k)\mathbb{P}(M_0 = x_0) \prod_{t=1}^{k} \mathbb{P}(M_t = x_t \mid M_{t-1} = x_{t-1}) \\ &= \mathbb{P}(M_{k+1} = x_{k+1} \mid M_k = x_k)\mathbb{P}(\mathrm{Traj}_M\restriction_{[k+1]-1} = x\restriction_{[k+1]-1}),\end{aligned}$$

so that whenever $\mathbb{P}(\mathrm{Traj}_M\restriction_{[k+1]-1} = s\restriction_{[k+1]-1}) > 0$, the conditional probabilities are well defined and (3.1) holds. This is precisely the Markov property. The fact that M has a finite state space holds since its state space is V, which is assumed to be finite. Moreover, Proposition 3.13 ensures that for every $t \in \mathbb{N}$ and $i, j \in V$,

$$\mathbb{P}(X_{t+1} = j \mid X_t = i) = \mathbb{P}(T_i X_t = j) = \mathbb{P}\big(X_t \in T_i^{-1}(j)\big),$$

whenever the conditional probability is defined. Since the last probability is precisely the measure of $T_i^{-1}(j)$ in $[0, 1]$, we have that it is independent of t, and that X is time-homogeneous. □

Note that given the chain specification (X, V, T, μ), it is possible to calculate the transition matrix and the initial probability of the Markov chain M defined by it. It is enough to inspect T and μ, respectively. Suppose that for every $i, j \in V$, the set $T_i^{-1}(j)$ is an interval. Proposition 3.13, then ensures

$$\begin{aligned}\mathbb{P}(M_1 = j \mid M_0 = i) &= \mathbb{P}(T_i X_1 = j) \\ &= \mathbb{P}(X_1 \in T_i^{-1}(j)) \\ &= |T_i^{-1}(j)|.\end{aligned}$$

Similarly, if for every $i \in V$, the set $\mu^{-1}(i)$ is an interval, then $\mathbb{P}(X_0 = i) = |\mu^{-1}(i)|$. These equalits hints at the role of Proposition 3.11 in the construction of Markov chains.

Theorem 3.15. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a chainable probability space. Let V be a finite set. Let $Y \colon \Omega \to V$ be a random variable, and let $P \in \mathbb{R}^{V \times V}$ be a stochastic matrix. Then there exists a time-homogeneous Markov chain M in $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V, and such that for every $i \in V$,

$$\mathbb{P}(M_0 = i) = \mathbb{P}(Y = i),$$

and for every i and j in V, and $t \in \mathbb{N}$,

$$\mathbb{P}(M_{t+1} = j \mid M_t = i) = P_{ij},$$

whenever the transition probability is defined.

Proof. Using Proposition 3.11 it is possible to ensure that there exists μ and T such that

1. For every $i \in V$, the set $\mu^{-1}(i)$ is and interval, and $|\mu^{-1}(i)| = \mathbb{P}(Y = i)$,
2. For every $i, j \in V$, the set $T_i^{-1}(j)$ is an interval, and $|T_i^{-1}(j)| = P_{ij}$.

Since $(\Omega, \mathcal{F}, \mathbb{P})$ is chainable, there exists a stochastic process X such that (X, V, T, μ) is a chain specification. The Markov chain M defined by it is precisely the object in the theorem statement, so that Proposition 3.14 completes the proof. □

Corollary 3.16. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a chainable probability space, and let $Y \colon \Omega \to V$ be a random variable. Then there exists a random walk on G such that for every $i \in V$,

$$\mathbb{P}(X_0 = i) = \mathbb{P}(Y = i)$$

Proof. Let P be the transition matrix of G. If we show that P^T is a stochastic matrix, then Theorem 3.15 finishes the proof. Let $i, j \in V$. Proposition 1.54 ensures that $D_G e_i = (\sum w(\delta(i)))\, e_i$. Since D_G is diagonal, this implies that

$$D_G^{-1} e_i = \frac{1}{\sum w(\delta(i))}.$$

Hence, we have that

$$P_{ij} = e_i^\mathsf{T} D_G^{-1} A_G e_j = (D_G^{-1} e_i)^\mathsf{T} A_G e_j = \frac{1}{\sum w(\delta(i))} (A_G)_{ij}.$$

Proposition 1.56 then ensures that

$$P_{ij}^\mathsf{T} = [ij \in E] \frac{w(ij)}{\sum w(\delta(i))}.$$

This implies that P^T is stochastic; every entry is nonnegative, and

$$e_i^\mathsf{T} P^\mathsf{T} \mathbb{1} = \frac{\sum w(\delta(i))}{\sum w(\delta(i))} = 1.$$

Theorem 3.15 then concludes the proof. □

3.3 Sequences of Vertices and Markov Chain Shifting

Let X be a time-homogeneous Markov chain with finite state space V set. Let $t \in \mathbb{N}$ be nonzero, and let $s \in V^{[t]-1}$. A simple inductive argument proves that

$$\mathbb{P}\Big(\mathrm{Traj}_X\restriction_{[t]-1} = s\Big) = \mathbb{P}(X_0 = s_0) \prod_{i=1}^{t-1} \mathbb{P}(X_i = s_i \mid X_{i-1} = s_{i-1}). \tag{3.17}$$

An interesting consequence of (3.17) is the fact that one can shift the focus from the probability space in which a Markov chain is defined to the set of sequences of elements in the state space of the chain. Let V

be a finite set. To work with this idea, we start by endowing $V^{\mathbb{N}}$ with an appropriate σ-algebra. A good reference for this construction is [12].

Let V be a finite set. Define a σ-algebra on $V^{\mathbb{N}}$ as $\sigma(\mathcal{O}_V)$, where

$$\mathcal{O}_V \coloneqq \bigcup_{t\in\mathbb{N}} \bigcup_{r\in V^{[t]}} \Big\{ s \in V^{\mathbb{N}} : s\restriction_{[t]} = r \Big\}.$$

Let X be a time-homogeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. The fact that Equation (3.2) holds for every $n \in \mathbb{N}$ and for every $t \in V^{[n]-1}$ can now be stated as the fact that for every event $E \in \mathcal{O}_V$, we have that $\{\mathrm{Traj}_X \in E\}_\Omega \in \mathcal{F}$. Theorem 1.73 then ensures that Traj_X is actually a measurable function with respect to $\mathcal{F}$ and $\sigma(\mathcal{O}_V)$. Moreover, define

$$\sigma[\mathrm{Traj}_X] \coloneqq \sigma(\{\, \mathrm{Traj}_X^{-1}(E) : E \in \mathcal{O}_V\}).$$

Since Traj_X is measurable, we have that $\sigma[\mathrm{Traj}_X] \subseteq \mathcal{F}$. This is helpful, since it is enough to show that a set is in $\sigma[\mathrm{Traj}_X]$ to ensure that it is measurable — and this can be quite easier. Moreover, some statements in this section will only apply to events in $\sigma[\mathrm{Traj}_X]$. Those are precisely the results obtained from the structure of $V^{\mathbb{N}}$.

Proposition 3.18. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space such that $\mathcal{F} = \sigma(\mathcal{O})$, for some $\mathcal{O} \subseteq \mathcal{P}(\Omega)$. Suppose $\mu\colon \mathcal{F} \to \mathbb{R}_+ \cup \{\infty\}$ is a probability measure on $(\Omega, \mathcal{F})$ such that $\mu(E) = \mathbb{P}(E)$ for every event $E \in \mathcal{O}$. Then $\mu = \mathbb{P}$.

Proof. Define

$$\mathcal{G} \coloneqq \{\, E \in \mathcal{F} : \mu(E) = \mathbb{P}(E)\}.$$

Since both $\mathbb{P}$ and μ are measures, we have that $\mathbb{P}(\varnothing) = 0 = \mu(\varnothing)$. Therefore, $\varnothing \in \mathcal{G}$. Moreover, suppose $E \in \mathcal{G}$. Then

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E) = 1 - \mu(E) = \mu(E^c).$$

Hence, $E^c \in \mathcal{G}$.

Let $(E_i)_{i\in\mathbb{N}}$ be a sequence of pairwise disjoint events in $\mathcal{G}$. Then

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} E_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(E_i) = \sum_{i\in\mathbb{N}} \mu(E_i) = \mu\left(\bigcup_{i\in\mathbb{N}} E_i\right).$$

This does not quite ensures that $\mathcal{G}$ is a σ-algebra, since the family of events had to be pairwise disjoint. Let then $(F_i)_{i\in\mathbb{N}}$ be any sequence of events in $\mathcal{G}$. Define, for every $i \in \mathbb{N}$,

$$E_i \coloneqq F_i \setminus \left(\bigcup_{k=0}^{i-1} F_k\right).$$

Note that the events $(E_i)_{i\in\mathbb{N}}$ are pairwise disjoint, and that $\bigcup_{i\in\mathbb{N}} F_i = \bigcup_{i\in\mathbb{N}} E_i$. Hence, the previous argument ensures $\bigcup_{i\in\mathbb{N}} F_i \in \mathcal{G}$. Therefore, $\mathcal{G}$ is a σ-algebra. Furthermore, the statement ensures that $\mathcal{O} \subseteq \mathcal{G}$. Theorem 1.72 then concludes the proof, since it implies that

$$\mathcal{F} = \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{G}) = \mathcal{G} \subseteq \mathcal{F}. \qquad \square$$

Proposition 3.19. Let X be a time-homoegeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. Let Y be a time-homogeneous Markov chain in a probability space $(\Omega', \mathcal{G}, \mathbb{P}_Y)$, with the same state space as X. Let $i \in V$. If X and Y have the same transition matrix, then for every event $E \in \sigma(\mathcal{O}_V)$,

$$\mathbb{P}(\mathrm{Traj}_X \in E \mid X_0 = i) = \mathbb{P}_Y(\mathrm{Traj}_Y \in E \mid Y_0 = i),$$

whenever both conditional probabilities are defined.

Proof. Suppose $i \in V$ is such that $\mathbb{P}(X_0 = i)$ and $\mathbb{P}_Y(Y_0 = i)$ are both positive. Define $\mu\colon \sigma(\mathcal{O}_V) \to [0,1]$ and $\nu\colon \sigma(\mathcal{O}_V) \to [0,1]$ as, for every $E \in \sigma(\mathcal{O}_V)$,

$$\mu(E) \coloneqq \mathbb{P}(\mathrm{Traj}_X \in E \mid X_0 = i),$$
$$\nu(E) \coloneqq \mathbb{P}_Y(\mathrm{Traj}_Y \in E \mid Y_0 = i).$$

Since Traj_X and Traj_Y are measurable, both μ and ν are probability measures on the measurable space $(V^{\mathbb{N}}, \sigma(\mathcal{O}_V))$. Proposition 3.18 reduces the statement into proving that for every $E \in \mathcal{O}_V$, it holds that $\mu(E) = \nu(E)$. Let $P \in \mathbb{R}^{V \times V}$ be the common transition matrix. Let then $n \in \mathbb{N}$, and $s \in V^{[n]-1}$. Equation (3.17) implies that

$$\mathbb{P}\Big(\mathrm{Traj}_X\!\restriction_{[n]-1} = s \,\Big|\, X_0 = i\Big) = [s_0 = i] \prod_{t=1}^{n-1} P_{s_t, s_{t-1}}.$$

Since Y has the same transition matrix, the same Equation (3.17) finishes the proof:

$$\mathbb{P}\Big(\mathrm{Traj}_X\!\restriction_{[n]-1} = s \,\Big|\, X_0 = i\Big) = [s_0 = i] \prod_{t=1}^{n-1} P_{s_t, s_{t-1}} = \mathbb{P}_Y\Big(\mathrm{Traj}_Y\!\restriction_{[n]-1} = s \,\Big|\, Y_0 = i\Big). \qquad \square$$

As a first application of the results and definitions just discussed, we proceed to define the random variables that will be used in our demonstrations.

Definition 3.20. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. Let $i \in V$. The *arrival time at i* is the function $\mathrm{Arr}_{X,i}\colon \Omega \to \mathbb{N} \cup \{\infty\}$ defined as

$$\mathrm{Arr}_{X,i}(\omega) \coloneqq \inf\{\, t \in \mathbb{N} : X_t(\omega) = i\},$$

for every $\omega \in \Omega$. If $\omega \in \Omega$ is such that $X_t(\omega) \neq i$ for every $t \in \mathbb{N}$, then $\mathrm{Arr}_{X,i}(\omega) = \infty$ by convention.

Proposition 3.21. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. For every $i \in V$,

(i) $\mathrm{Arr}_{X,i}$ is a random variable, and

(ii) for every $t \in \mathbb{N}$ we have that $\{\mathrm{Arr}_{X,i} = t\}_\Omega \in \sigma[\mathrm{Traj}_X]$.

Proof. Note that (ii) implies (i), since $\sigma[\mathrm{Traj}_X] \subseteq \mathcal{F}$ and

$$\{\mathrm{Arr}_{X,i} = \infty\}_\Omega = \Omega \setminus \left(\bigcup_{t \in \mathbb{N}} \{\mathrm{Arr}_{X,i} = t\}_\Omega \right).$$

Let then $T \in \mathbb{N}$. Define

$$\mathcal{S}_{i,T} \coloneqq \{\, s \in V^{[T+1]-1} : s_T = i, s_t \neq i \quad \forall t \in [T] - 1\}.$$

For a given $\omega \in \Omega$, it holds that $\mathrm{Arr}_{X,i}(\omega) = T$ if and only if $X_T(\omega) = i$ and, for every $t \in \mathbb{N}$ with $t < T$ we have that $X_t(\omega) \neq i$. Therefore, $\{\mathrm{Arr}_{X,i} = T\}_\Omega = \Big\{\mathrm{Traj}_X\!\restriction_{[T+1]-1} \in \mathcal{S}_{i,T}\Big\}_\Omega$, and (ii) holds. $\square$

Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. If all the vertices have a finite arrival time, they have to be distinct, since the Markov chain can only have one value for each $\omega \in \Omega$ and $t \in \mathbb{N}$. This modest remark is actually quite important in the following definition.

Definition 3.22. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. The *last vertex* is the function $\mathrm{Last}_X\colon \Omega \to V \cup \{\bot\}$ defined as, for every $\omega \in \Omega$,

$$\mathrm{Last}_X(\omega) \coloneqq \begin{cases} \bot, & \text{if } \max_{i \in V} \mathrm{Arr}_{X,i}(\omega) = \infty, \\ \arg\max_{i \in V} \mathrm{Arr}_{X,i}(\omega), & \text{otherwise,} \end{cases}$$

where $\arg\max_{i \in V} \mathrm{Arr}_{X,i}(\omega)$ denotes the vertex $i \in V$ such that $\mathrm{Arr}_{X,i}(\omega)$ is maximum.

Proposition 3.23. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. Then

(i) Last_X is a random variable, and

(ii) for every $i \in V$ we have that $\{\mathrm{Last}_X = i\}_\Omega \in \sigma[\mathrm{Traj}_X]$.

Proof. Note that (ii) implies (i), since $\sigma[\mathrm{Traj}_X] \subseteq \mathcal{F}$ and

$$\{\mathrm{Last}_X = \bot\}_\Omega = \Omega \setminus \left(\bigcup_{i \in V} \{\mathrm{Last}_X = i\}_\Omega \right).$$

For every $i \in V$ and $T \in \mathbb{N}$, it is clear that

$$\{\mathrm{Arr}_{X,i} < T\}_\Omega = \bigcup_{t=0}^{T-1} \{\mathrm{Arr}_{X,i} = t\}_\Omega.$$

Proposition 3.21 ensures that every event in the RHS is in $\sigma[\mathrm{Traj}_X]$. Hence $\{\mathrm{Arr}_{X,i} < t\}_\Omega \in \sigma[\mathrm{Traj}_X]$. This finishes the proof, since for every $i \in V$,

$$\{\mathrm{Last}_X = i\}_\Omega = \bigcup_{t \in \mathbb{N}} \left(\{\mathrm{Arr}_{X,i} = t\}_\Omega \cap \bigcap_{j \in V \setminus \{i\}} \{\mathrm{Arr}_{X,j} < t\}_\Omega \right). \qquad \square$$

Definition 3.24. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. The *cover time* is the function $\mathrm{Cov}_X \colon \Omega \to \mathbb{N} \cup \{\infty\}$ defined as

$$\mathrm{Cov}_X(\omega) \coloneqq \max_{i \in V} \mathrm{Arr}_{X,i}(\omega),$$

for every $\omega \in \Omega$.

Proposition 3.25. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. Then

(i) Cov_X is a random variable, and

(ii) for every $t \in \mathbb{N}$ we have that $\{\mathrm{Cov}_X = t\}_\Omega \in \sigma[\mathrm{Traj}_X]$.

Proof. Once again, note that (ii) implies (i), since

$$\{\mathrm{Cov}_X = \infty\}_\Omega = \bigcup_{i \in V} \{\mathrm{Arr}_{X,i} = \infty\}_\Omega,$$

so that $\{\mathrm{Cov}_X = \infty\}_\Omega$ is measurable. Proposition 3.21 and Proposition 3.23 then conclude the proof, since

$$\{\mathrm{Cov}_X = t\}_\Omega = \bigcup_{i \in V} \left(\{\mathrm{Last}_X = i\}_\Omega \cap \{\mathrm{Arr}_{X,i} = t\}_\Omega \right). \qquad \square$$

We now focus on a second application of Proposition 3.18. Too often, arguments about time-homogeneous Markov chains work with the idea that, if you only consider a random walk from a specific point forward, the random walk is still the same, only with distinct initial distribution. This kind of reasoning usually produces recurrence relations involving random variables being considered, which can then produce further results. Unfortunately, these arguments tend to be as imprecise as they are useful. We proceed in establishing a solid language to work with this meaningful idea.

Definition 3.26. Let X be a Markov chain. The *shift of X* is the Markov chain Y defined as

$$Y_i \coloneqq X_{i+1}.$$

There are several remarks in place. First, the shift of random walk just "drops" its first random variable. This can easily be fixed via inductive arguments, by defining the nth shift. We do not do so, however, because this generality will not be necessary.

Now, let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. Let Y be the shift of X. By definition, Y is in the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and it has the same state space as X. Moreover, if X is time-homogeneous, then so is Y. Furthermore, for every $i, j \in V$

$$\mathbb{P}(Y_1 = j \mid Y_0 = i) = \mathbb{P}(X_2 = j \mid X_1 = i) = \mathbb{P}(X_1 = j \mid X_0 = i).$$

Hence,both Markov chains have the same transition matrix. This implies that both Markov chains have the same transition digraph and the same stationary distributions. Hence, if X is irreducible, then so is Y, and if X is time-reversible, then so is Y. Moreover, if X is a random walk on a graph G, then Y also is a random walk on G.

Proposition 3.27. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite state space V. Let Y be the shift of X. Then for every $\omega \in \Omega$, and for every $i \in V \setminus \{X_0(\omega)\}$,

$$\mathrm{Arr}_{Y,i}(\omega) = \mathrm{Arr}_{X,i}(\omega).$$

Proof. Let $\omega \in \Omega$, and $i \in V \setminus \{X_0(\omega)\}$. Then

$$\begin{aligned}\mathrm{Arr}_{Y,i}(\omega) &= \inf\{\, t \in \mathbb{N} : Y_t(\omega) = i\} \\ &= \inf\{\, t \in \mathbb{N} : X_{t+1}(\omega) = i\} \\ &= \inf\{\, t - 1 \in \mathbb{N} : X_t(\omega) = i\} \\ &= \inf\{\, t \in \mathbb{N} \setminus \{0\} : X_t(\omega) = i\} - 1 \\ &= \mathrm{Arr}_{X,i}(\omega) - 1.\end{aligned}$$

Note that the last equality relies on the fact that the state $X_0(\omega) \in V$ is the only one with arrival time 0. □

Proposition 3.28. Let X be a Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. Let $i, j \in V$. Let Y be the shift of X. For every event $E \in \sigma[\mathrm{Traj}_Y]$,

$$\mathbb{P}(E \mid Y_0 = j, X_0 = i) = \mathbb{P}(E \mid Y_0 = j),$$

whenever both conditional probabilities are defined.

Proof. Suppose $i \in V$ and $j \in V$ are such that $\mathbb{P}(Y_0 = j, X_0 = i) > 0$, so that both conditional probabilities are defined. Define $\mu\colon \sigma(\mathcal{O}_V) \to [0,1]$ and $\nu\colon \sigma(\mathcal{O}_V) \to [0,1]$ as, for every $E \in \sigma(\mathcal{O}_V)$,

$$\begin{aligned}\mu(E) &\coloneqq \mathbb{P}(\mathrm{Traj}_Y \in E \mid Y_0 = j, X_0 = i), \\ \nu(E) &\coloneqq \mathbb{P}(\mathrm{Traj}_Y \in E \mid Y_0 = j).\end{aligned}$$

Since Traj_Y is measurable, both ν and μ are probability measures on the measurable space $(V^{\mathbb{N}}, \sigma(\mathcal{O}_V))$. Proposition 3.18 reduces the statement into proving that for every $E \in \mathcal{O}_V$, it holds that $\mu(E) = \nu(E)$. However, the Markov property (3.1) is precisely this statement, since it implies that for every $n \in \mathbb{N}$ and $s \in V^{[n]-1}$,

$$\mathbb{P}\Big(\mathrm{Traj}_Y\restriction_{[n]-1} = s \,\Big|\, Y_0 = j, X_0 = i\Big) = \mathbb{P}\Big(\mathrm{Traj}_Y\restriction_{[n]-1} = s \,\Big|\, Y_0 = j\Big). \qquad \square$$

We close this section with an argument using the shift of a random walk. It serves both as an application of the theory just developed, and as a preview of the proof of correctness of the Aldous-Broder algorithm.

Proposition 3.29. Let X be a time-homogeneous Markov chain in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. If X is irreducible, then for every $r \in V$,

$$\mathbb{P}(\mathrm{Arr}_{X,r} = \infty) = 0.$$

Moreover, $\mathbb{P}(\mathrm{Cov}_X = \infty) = 0$.

Proof. Note that

$$\{\mathrm{Cov}_X = \infty\}_\Omega = \bigcup_{r\in V} \{\mathrm{Arr}_{X,r} = \infty\}_\Omega.$$

Suffices then to show that for every $r \in V$, we have that $\mathbb{P}(\mathrm{Arr}_{X,r} = \infty) = 0$. We first prove this statement assuming that X is a Markov chain such that, for every $i \in V$, it holds that $\mathbb{P}(X_0 = i) > 0$.

Let $r \in V$. For every $i \in V \setminus \{r\}$, we have that

$$\begin{aligned}\{\mathrm{Arr}_{X,r} = \infty, X_0 = i\}_\Omega &= \bigcup_{j\in V} \{\mathrm{Arr}_{X,r} = \infty, Y_0 = j, X_0 = i\}_\Omega \\ &= \bigcup_{j\in V} \{\mathrm{Arr}_{Y,r} = \infty, Y_0 = j, X_0 = i\}_\Omega.\end{aligned}$$

Moreover, Proposition 3.21 ensures that both Proposition 3.19 and Proposition 3.28 apply. Therefore, for every $i \in V \setminus \{r\}$,

$$\begin{aligned}\mathbb{P}(\mathrm{Arr}_{X,r} = \infty \mid X_0 = i) &= \sum_{j\in V} \mathbb{P}(\mathrm{Arr}_{Y,r} = \infty, Y_0 = j \mid X_0 = i) \\ &= \sum_{j\in V} \mathbb{P}(\mathrm{Arr}_{Y,r} = \infty \mid Y_0 = j, X_0 = i)\mathbb{P}(Y_0 = j \mid X_0 = i) \\ &= \sum_{j\in V} \mathbb{P}(\mathrm{Arr}_{Y,r} = \infty \mid Y_0 = j)\mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \sum_{j\in V} \mathbb{P}(\mathrm{Arr}_{X,r} = \infty \mid X_0 = j)\mathbb{P}(X_1 = j \mid X_0 = i).\end{aligned}$$

Denote then $f_r(i) \coloneqq \mathbb{P}(\mathrm{Arr}_{X,r} = \infty \mid X_0 = i)$. We have just proved that function f_r is harmonic, with respect to the transition digraph of X, at every vertex in $V \setminus \{r\}$. Since this digraph is strongly connected, Proposition 1.62 ensures that f_r is constant. Furthermore, since $f_r(r) = 0$, we have that $f_r = 0$. Therefore, for every $r \in V$, we have that $\mathbb{P}(\mathrm{Arr}_{X,r} = \infty) = 0$.

Let Y be any time-homogeneous Markov chain with finite state space V. Theorem 3.15 ensures that there exists a Markov chain X with the same transition matrix as Y and such that for every $i \in V$ we have that $\mathbb{P}(X_0 = i) = 1/|V|$. We have just proved that for every $r \in V$, we have that $\mathbb{P}(\mathrm{Arr}_{X,r} = \infty \mid X_0 = i) = 0$. Proposition 3.21 and Proposition 3.19 then finish the proof, since for every $r \in V$,

$$\begin{aligned}\mathbb{P}(\mathrm{Arr}_{Y,r} = \infty) &= \sum_{i\in V} [\mathbb{P}(Y_0 = i) > 0]\, \mathbb{P}(\mathrm{Arr}_{Y,r} = \infty \mid Y_0 = i)\mathbb{P}(Y_0 = i) \\ &= \sum_{i\in V} [\mathbb{P}(Y_0 = i) > 0]\, \mathbb{P}(\mathrm{Arr}_{X,r} = \infty \mid X_0 = i)\mathbb{P}(Y_0 = i) \\ &= 0.\end{aligned}$$

□

Proposition 3.29 has the important consequence of ensuring that our crass definition of expected value of random variables apply to both the arrival and cover time. Let X be a time-homogeneous, irreducible Markov chain, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a finite state space V. Proposition 3.29 then ensures that $\mathbb{P}(\mathrm{Cov}_X = \infty) = 0$. Therefore, $1 = \sum_{t\in\mathbb{N}} \mathbb{P}(\mathrm{Cov}_X = t)$. Hence, according to Definition 1.75, we have

$$\mathbb{E}[\mathrm{Cov}_X] = \sum_{t\in\mathbb{N}} t\, \mathbb{P}(\mathrm{Cov}_X = t). \tag{3.30}$$

Likewise, let $i \in V$. Proposition 3.29 ensures that $\mathbb{P}(\mathrm{Arr}_{X,i} = \infty) = 0$. Therefore, $1 = \sum_{t\in\mathbb{N}} \mathbb{P}(\mathrm{Arr}_{X,i} = t)$. Hence, according to Definition 1.75, we have

$$\mathbb{E}[\mathrm{Arr}_{X,i}] = \sum_{t\in\mathbb{N}} t\, \mathbb{P}(\mathrm{Arr}_{X,i} = t). \tag{3.31}$$

Chapter 4

The Aldous-Broder Algorithm

This chapter is about the result first proved by Aldous [1] and Broder [5] — hence the section name. The algorithm described is well defined on every Markov chain. However, further hypothesis are needed to ensure its correctness. Broder requires the Markov chain to be irreducible and time-reversible, whereas Aldous define the algorithm for random walks on graphs. As was already established in the previous chapter, both approaches are equivalent; to work with irreducible, time-reversible Markov chains is to work with a random walk in a connected weighted graph.

Since only connected graphs have spanning trees, every graph mentioned in a statement in this chapter will be connected. A convenient consequence is that every random walk we deal with is an irreducible Markov chain. Hence, we can (and will) simply use Equation (3.30) and Equation (3.31) to calculate the expected values of the random variables we are working with.

Moreover, we assume the graphs we are dealing are simple. To remove loops is clearly not a limitation, since loops cannot be in a spanning tree. The case for parallel edges is more interesting, and for such reason we proceed to sketch how one could use an algorithm that samples in simple graphs to sample in graphs with parallel edges.

Suppose $G = (V, E, \psi, w)$ is a weighted graph, with $w \in \mathbb{R}^E_{++}$ and $e, f \in E$ are such that $\psi(e) = \psi(f)$. For simplicity, suppose e and f are the only such pair. Define the graph $H := (V, E - f, \psi\restriction_{E-f}, w')$, where w' is equal to w in every edge but e, where it is defined as $w'(e) := w(e) + w(f)$. The graph H is simple. Suppose we sampled a spanning tree T in G. If $e \neq T$, we just output T as a spanning tree of G. However, if $e \in T$, we can output T with probability $w(e)/(w(e) + w(f))$, and $T - e + f$ with probability $w(f)/(w(e) + w(f))$. It is then possible to use fact that the algorithm was correct for G to conclude that the algorithm is correct for H.

4.1 The Algorithm

Definition 4.1. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *Aldous-Broder algorithm* is the function $\mathcal{A}(X)\colon \Omega \to \mathcal{P}(E) \cup \{\perp\}$ defined, for every $\omega \in \Omega$, as

$$\mathcal{A}(X)(\omega) := \begin{cases} \perp, & \text{if } \operatorname{Cov}_X(\omega) = \infty, \\ \bigcup_{i \in V \setminus \{X_0(\omega)\}} \{\{X_{\operatorname{Arr}_{X,i}(\omega)-1}, X_{\operatorname{Arr}_{X,i}(\omega)}\}\}, & \text{otherwise.} \end{cases}$$

Proposition 4.2. Let G be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For every $\omega \in \Omega$, if $\mathcal{A}(X)(\omega) \neq \perp$, then $\mathcal{A}(X)(\omega) \in \mathcal{T}_G$.

Proof. Let $G = (V, E, w)$. Let $\omega \in \Omega$ be such that $\mathcal{A}(X)(\omega) \neq \perp$. The definition of the algorithm then ensures that $\operatorname{Cov}_X(\omega) < \infty$. Therefore, $\operatorname{Arr}_{X,i}(\omega) < \infty$ for every $i \in V(G)$. Hence, the set

$$F := \mathcal{A}(X)(\omega) = \bigcup_{i \in V \setminus \{X_0(\omega)\}} \{\{X_{\operatorname{Arr}_{X,i}(\omega)-1}, X_{\operatorname{Arr}_{X,i}(\omega)}\}\}$$

has $|V| - 1$ edges in it. Suffices to show that the graph $G[F]$ is connected.

Set $i \coloneqq X_0(\omega)$. We proceed to show that every vertex in V is connected to i. This can be done by induction on $\mathrm{Arr}_{X,j}(\omega)$, for every vertex j. If $\mathrm{Arr}_{X,j}(\omega) = 0$, then $j = i$ and the thesis holds. Otherwise, let $j \in V$ be such that $t \coloneqq \mathrm{Arr}_{X,j}(\omega)$ is nonzero. Set $k \coloneqq X_{t-1}(\omega)$. Since $\mathrm{Arr}_{X,k}(\omega) \leq t - 1 < t$, the induction hypothesis ensures that k is connected to i. Moreover, $kj \in F$. Hence, j is connected to i. □

Definition 4.3. Let G be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let $T \in \mathcal{T}_G$. Define $\mathcal{S}_{T,0} \coloneqq \varnothing$. Moreover, for every nonzero $t \in \mathbb{N}$, define $\mathcal{S}_{T,t}$ as the set of sequences $s \in V^{[t]-1}$ such that for every random walk X on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, it holds that

$$\left\{\mathrm{Traj}_X\restriction_{[t]-1} = s\right\}_\Omega \subseteq \{\mathcal{A}(X) = T, \mathrm{Cov}_X = t - 1\}_\Omega.$$

Proposition 4.4. Let G be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For every $T \in \mathcal{T}_G$,

$$\{\mathcal{A}(X) = T\}_\Omega = \bigcup_{t \in \mathbb{N}} \left\{\mathrm{Traj}_X\restriction_{[t]-1} \in \mathcal{S}_{T,t}\right\}_\Omega.$$

Proof. Note that, by definition, $\{\mathcal{A}(X) = T\}_\Omega \subseteq \{\mathrm{Cov}_X < \infty\}_\Omega$. Hence

$$\{\mathcal{A}(X) = T\}_\Omega = \{\mathcal{A}(X) = T, \mathrm{Cov}_X < \infty\}_\Omega = \bigcup_{t \in \mathbb{N}} \{\mathcal{A}(X) = T, \mathrm{Cov}_X = t\}_\Omega.$$

Using this equality and the definition of $\mathcal{S}_{T,t}$, we have that

$$\bigcup_{t \in \mathbb{N}} \left\{\mathrm{Traj}_X\restriction_{[t]-1} \in \mathcal{S}_{T,t}\right\}_\Omega = \bigcup_{t=1}^{\infty} \left\{\mathrm{Traj}_X\restriction_{[t]-1} \in \mathcal{S}_{T,t}\right\}_\Omega \subseteq \bigcup_{t=1}^{\infty} \{\mathcal{A}(X) = T, \mathrm{Cov}_X = t - 1\}_\Omega = \{\mathcal{A}(X) = T\}_\Omega.$$

It remains to show that $\{\mathcal{A}(X) = T\}_\Omega \subseteq \bigcup_{t \in \mathbb{N}} \left\{\mathrm{Traj}_X\restriction_{[t]-1} \in \mathcal{S}_{T,t}\right\}_\Omega$. Let then $\omega \in \{\mathcal{A}(X) = T\}_\Omega$. Since $\mathcal{A}(X)(\omega) \neq \bot$, it holds that $\mathrm{Cov}_X(\omega) < \infty$. Define $t \coloneqq \mathrm{Cov}_X(\omega) + 1$. Let $s \in V^{[t]-1}$ be defined as $s_i \coloneqq X_i(\omega)$, for every $i \in [t] - 1$. Since $\omega \in \left\{\mathrm{Traj}_X\restriction_{[t]-1} = s\right\}_\Omega$, suffices to show that $s \in \mathcal{S}_{T,t}$. Suppose then that Y is a random walk on G, in a probability space $(\Omega', \mathcal{G}, \mathbb{P}_Y)$. We wish to show that

$$\left\{\mathrm{Traj}_Y\restriction_{[t]-1} = s\right\}_{\Omega'} \subseteq \{\mathcal{A}(Y) = T, \mathrm{Cov}_Y = t - 1\}_{\Omega'}.$$

Suppose $\omega' \in \left\{\mathrm{Traj}_Y\restriction_{[t]-1} = s\right\}_{\Omega'}$. Note that this implies that $Y_k(\omega') = s_k = X_k(\omega)$ for every $k \in [t] - 1$. Therefore, for every $i \in V$, we have that $\mathrm{Arr}_{X,i}(\omega) = \mathrm{Arr}_{Y,i}(\omega')$. Moreover, $\mathrm{Cov}_Y(\omega') = \mathrm{Cov}_X(\omega) = t - 1$. Hence

$$\omega' \in \{\mathcal{A}(Y) = T, \mathrm{Cov}_Y = t - 1\}_{\Omega'}.$$

Since Y was an arbitrary random walk on G, it follows that $s \in \mathcal{S}_{T,t}$. □

Corollary 4.5. Let G be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G. The Aldous-Broder algorithm $\mathcal{A}(X)$ is a random variable.

Proof. Suppose X is a random walk in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Proposition 4.4 ensures that for every $T \in \mathcal{T}_G$, the event $\{\mathcal{A}(X) = T\}_\Omega \in \sigma[\mathrm{Traj}_X]$. The proof is done, since

$$\{\mathcal{A}(X) = \bot\}_\Omega = \Omega \setminus \{\mathcal{A}(X) \in \mathcal{T}_G\}_\Omega,$$

so that $\{\mathcal{A}(X) = \bot\}_\Omega \in \sigma[\mathrm{Traj}_X]$. □

Proposition 4.6. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let Y be the shift of X. For every $T \in \mathcal{T}_G$ and for every $i \in V$,

$$\{\mathcal{A}(X) = T, X_0 = i, \mathcal{A}(Y) \neq \bot\}_\Omega = \bigcup_{j \in \Gamma_T(i)} \bigcup_{k \in C_j} \{\mathcal{A}(Y) = T - ij + ki, Y_0 = j, X_0 = i\}_\Omega,$$

where C_j is the component of j in $G[T] - i$ intersected with $\Gamma(i)$.

Proof. Let $\omega \in \{\mathcal{A}(X) = T, X_0 = i, \mathcal{A}(Y) \neq \bot\}_\Omega$. The fact that $\mathcal{A}(Y)(\omega) \neq \bot$ implies $\mathrm{Arr}_{Y,i}(\omega) < \infty$. Set $t \coloneqq \mathrm{Arr}_{Y,i}(\omega)$ and $k \coloneqq Y_{t-1}(\omega)$, so that $ki \in \mathcal{A}(Y)(\omega)$. Set $j \coloneqq X_1(\omega) = Y_0(\omega)$. Proposition 3.27 ensures

$$\begin{aligned}
\mathcal{A}(Y)(\omega) &= \bigcup_{\ell \in V \setminus \{j\}} \big\{\{Y_{\mathrm{Arr}_{Y,\ell}(\omega)-1}, Y_{\mathrm{Arr}_{Y,\ell}(\omega)}\}\big\} \\
&= \{ki\} \cup \bigcup_{\ell \in V \setminus \{i,j\}} \big\{\{Y_{\mathrm{Arr}_{Y,\ell}(\omega)-1}, Y_{\mathrm{Arr}_{Y,\ell}(\omega)}\}\big\} \\
&= \{ki\} \cup \bigcup_{\ell \in V \setminus \{i,j\}} \big\{\{Y_{\mathrm{Arr}_{X,\ell}(\omega)-2}, Y_{\mathrm{Arr}_{X,\ell}(\omega)-1}\}\big\} \\
&= \{ki\} \cup \bigcup_{\ell \in V \setminus \{i,j\}} \big\{\{X_{\mathrm{Arr}_{X,\ell}(\omega)-1}, X_{\mathrm{Arr}_{X,\ell}(\omega)}\}\big\},
\end{aligned}$$

so that $\mathcal{A}(Y)(\omega) = T - ij + ki$. This ensures that the set in the LHS of the statement is a subset of the one in the RHS.

Let $j \in \Gamma_T(i)$, and $k \in C_j$. Suppose then that $\omega \in \{\mathcal{A}(Y) = T - ij + ki, Y_0 = j, X_0 = i\}_\Omega$. Since $\mathcal{A}(Y)(\omega)$ is a tree, we have that $\omega \in \{\mathcal{A}(Y) \neq \bot\}_\Omega$. Suffices to show that $\mathcal{A}(X)(\omega) = T$. Proposition 3.27 ensures that

$$\begin{aligned}
\mathcal{A}(X)(\omega) &= \bigcup_{\ell \in V \setminus \{i\}} \big\{\{X_{\mathrm{Arr}_{X,\ell}(\omega)-1}, X_{\mathrm{Arr}_{X,\ell}(\omega)}\}\big\} \\
&= \{ij\} \cup \bigcup_{\ell \in V \setminus \{i,j\}} \big\{\{X_{\mathrm{Arr}_{X,\ell}(\omega)-1}, X_{\mathrm{Arr}_{X,\ell}(\omega)}\}\big\} \\
&= \{ij\} \cup \bigcup_{\ell \in V \setminus \{i,j\}} \big\{\{Y_{\mathrm{Arr}_{Y,\ell}(\omega)-1}, Y_{\mathrm{Arr}_{Y,\ell}(\omega)}\}\big\}.
\end{aligned}$$

Therefore $\mathcal{A}(X)(\omega) = (T - ij + ki) - ki + ij = T$, and the proof is finished. □

Proposition 4.7. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $i \in V$ be such that $\mathbb{P}(X_0 = i) > 0$. For every $T \in \mathcal{T}_G$, denote by j_k be the first vertex in the ik-path in $G[T]$. Then

$$\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i) = \sum_{k \in \Gamma(i)} \mathbb{P}(X_1 = j_k \mid X_0 = i)\mathbb{P}(\mathcal{A}(X) = T - ij_k + ik \mid X_0 = j_k).$$

Proof. Let Y be the shift of X. Since X is irreducible, so is Y. Proposition 3.29 then ensures that $\mathbb{P}(\mathcal{A}(Y) = \bot) = 0$. Hence $\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i) = \mathbb{P}(\mathcal{A}(X) = T, \mathcal{A}(Y) \neq \bot \mid X_0 = i)$. Moreover, since $\mathbb{P}(X_0 = i) > 0$, for every $j \in \Gamma_T(i)$ we have that

$$\mathbb{P}(Y_0 = j) \geq \mathbb{P}(Y_0 = j, X_0 = i) = \mathbb{P}(Y_0 = j \mid X_0 = i)\mathbb{P}(X_0 = i) = \frac{w(ij)}{\sum w(\delta(i))}\mathbb{P}(X_0 = i) > 0.$$

Hence, for every $j \in \Gamma_T(i)$, the conditional probabilities $\mathbb{P}(\cdot \mid Y_0 = j)$ and $\mathbb{P}(\cdot \mid Y_0 = j, X_0 = i)$ are defined. Proposition 4.6 and Proposition 3.28 then ensure

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i) &= \sum_{j \in \Gamma_T(i)} \sum_{k \in C_j} \mathbb{P}(\mathcal{A}(Y) = T - ij + ik, Y_0 = j \mid X_0 = i) \\
&= \sum_{j \in \Gamma_T(i)} \sum_{k \in C_j} \mathbb{P}(\mathcal{A}(Y) = T - ij + ik \mid Y_0 = j, X_0 = i)\mathbb{P}(Y_0 = j \mid X_0 = i) \\
&= \sum_{j \in \Gamma_T(i)} \sum_{k \in C_j} \mathbb{P}(\mathcal{A}(Y) = T - ij + ik \mid Y_0 = j)\mathbb{P}(X_1 = j \mid X_0 = i),
\end{aligned}$$

where C_j is the component of j in $G[T] - i$ intersected with $\Gamma(i)$.

To finish the proof, remains only to change the summation order. Since G is simple, $\Gamma(i) \subseteq V \setminus \{i\}$. Hence, every $k \in \Gamma(i)$ belongs to a single component in $G[T] - i$. Therefore, for every $k \in \Gamma(i)$ there is a single $j \in \Gamma_T(i)$ such that (j,k) appear as indices in the double summation. Moreover, since k belongs to the component of j in $G[T] - i$, there is a unique ik-path in $G[T]$, and $ij \in T$, we have that j is the first vertex in such path. Hence

$$\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i) = \sum_{k \in \Gamma(i)} \mathbb{P}(X_1 = j_k \mid X_0 = i)\mathbb{P}(\mathcal{A}(Y) = T - ij_k + ik \mid Y_0 = j_k),$$

and Proposition 3.19 finishes the proof. □

Definition 4.8. Let $G = (V, E)$ be a simple, connected graph. Define $\mathcal{H}(G) \coloneqq (\mathcal{T}_G \times V, F)$ to be such that for every $T \in \mathcal{T}_G$ and $i \in V$,

$$\Gamma_{\mathcal{H}(G)}(T, i) \coloneqq \bigcup_{k \in \Gamma(i)} \{(T - ij_k + ik, j_k)\},$$

where j_k is the first vertex in the ik-path in T.

Let G be a simple, connected graph. It will be an important step to prove the correctness of the Aldous-Broder algorithm to show that $\mathcal{H}(G)$ is connected. The next 3 propositions prove that this is the case. The main proof will follow easily from Proposition 4.9 and Proposition 4.10, both propositions which will be proved via equivalent techniques. However, it is remarkable how the first is as simple as an inductive argument gets, and the second is a full of parameters minimal counterexample demonstration.

Proposition 4.9. Let $G = (V, E)$ be a simple, connected graph. For every $T \in \mathcal{T}_G$ and $i, j \in V$, we have that (T, i) and (T, j) are connected in $\mathcal{H}(G)$.

Proof. Since T is a spanning tree, for every vertices $i, j \in V$ there exists an ij-walk in $G[T]$. Let $(i, e_1, \dots, e_\ell, j)$ be this walk. We proceed by induction on ℓ, the length of the walk.

If $\ell = 0$, then $i = j$, and there is nothing to prove. Assume $\ell > 0$, and let k be the first vertex in the ij-walk under consideration. The induction hypothesis implies that (T, k) and (T, j) are connected, since the kj-walk $(k, e_2, \dots, j)$ has length $\ell - 1$. Moreover, since k is the first vertex in the unique ik-walk in $G[T]$, we have that (T, i) is adjacent to

$$(T - ik + ik, k) = (T, k)$$

in $\mathcal{H}(G)$, and the proof is finished. □

Proposition 4.10. Let $G = (V, E, \psi)$ be a simple, connected graph. Let $r \in V$, let $T \in \mathcal{T}_G$, and $e_0 \in E \setminus T$. Denote by $ik \coloneqq \psi(e_0)$. Then for every edge e in the only ik-path in $G[T]$, the pair (T, r) is connected to $(T - e + e_0, r)$ in $\mathcal{H}(G)$.

Proof. For every $T \in \mathcal{T}_G$ and for every $i, j \in V$, there is a unique ij-path in $G[T]$. Denote by $\ell_T(i, j)$ the length of this walk. For every $T \in \mathcal{T}_G$ and $e, e_0 \in E$, define

$$d_T(e, e_0) \coloneqq \min\{\, \ell_T(i, j) : i \in \psi(e), j \in \psi(e_0)\}.$$

Suppose that the graph $G = (V, E, \psi)$, the vertex $r \in V$, the spanning tree $T \in \mathcal{T}_G$, and the edges $e \in T$ and $e_0 \in E \setminus T$ are a counterexample that minimizes $d_T(e, e_0)$. We first prove that $d_T(e, e_0) > 0$.

If $d_T(e, e_0) = 0$, then $|\psi(e) \cap \psi(e_0)| > 0$. Let $i \in \psi(e) \cap \psi(e_0)$, and let $j, k \in V$ be such that $\psi(e_0) = ik$ and $\psi(e) = ij$. Proposition 4.9 ensures that (T, r) is connected to (T, i), and since

$$T - e + e_0 = T - ij + ik,$$

we have that (T, i) and $(T - e + e_0, j)$ are adjacent in $\mathcal{H}(G)$. Proposition 4.9 then ensures that $(T - e + e_0, j)$ and $(T - e + e_0, r)$ are connected in $\mathcal{H}(G)$, which contradicts the fact that we were dealing with a counterexample.

Set $t \coloneqq d_T(e, e_0)$. We have just proved that $t > 0$. Without loss of generality, assume that $i, k \in V$ are such that $\ell_T(i, k) = d_T(e, e_0)$. Let $(i, e_1, j, \dots, e_t, k)$ be the only ik-walk in $G[T]$. Proposition 4.9 ensures that (T, r) is connected to (T, i), and the definition of $\mathcal{H}(G)$ implies that (T, i) is adjacent to $(T - e_1 + e_0, j)$. Moreover, note that

$$d_{T - e_1 + e_0}(e_1, e) = t - 1,$$

and that $e_1 \in E \setminus (T - e_1 + e_0)$. Hence, we have that $(T - e_1 + e_0, r)$ is connected to $(T - e + e_0, r)$. Therefore, we have that (T, r) is connected to (T, i), which is adjacent to $(T - e_1 + e_0, j)$, which is then connected to $(T - e + e_0, r)$. In other words, (T, r) is connected to $(T - e + e_0, r)$. This contradicts the choice of parameters as a counterexample. □

Proposition 4.11. Let G be a simple, connected graph. The graph $\mathcal{H}(G)$ is connected.

Proof. Let $G = (V, E, \psi)$. For every $S, T \in \mathcal{T}_G$, define

$$d(S,T) \coloneqq |V| - 1 - |S \cap T|\,.$$

Let $r \in V$. We proceed to prove that for every $S, T \in \mathcal{T}_G$, the pairs (T, r) and (S, r) are connected in $\mathcal{H}(G)$. This proof will be by induction on $d(S, T)$. Note that this and Proposition 4.9 imply the statement.

If $d(S, T) = 0$, then $S = T$, and there is nothing to prove. Let then T and S be spanning trees such that $d(S, T) > 0$. Since $d(S, T) \neq 0$, there exists $e_0 \in S \setminus T$. Set $ik \coloneqq \psi(e_0)$. Since G is simple, we have that $i \neq k$. Moreover, since S is a tree, it is acyclic. Hence, there exists an edge e in the ik-walk in $G[T]$ that is not in S. In other words, there exists an edge e in the ik-walk in $G[T]$ such that $e \in T \setminus S$. Proposition 4.10 then ensures that (T, r) is connected to $(T - e + e_0, r)$. But since $e_0 \in S \setminus T$ and $e \in T \setminus S$, we have that $d(T - e + e_0, S) = d(T, S) - 1$. The induction hypothesis then finishes the proof. □

Theorem 4.12. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For every $T \in \mathcal{T}_G$,

$$\mathbb{P}(\mathcal{A}(X) = T) = \frac{\prod w(T)}{\Phi(G)}.$$

Proof. As we did in Proposition 3.29, we first prove the statement assuming that for every $i \in V$ we have that $\mathbb{P}(X_0 = i) > 0$. Define the function $Z \colon \mathcal{T}_G \times V \to \mathbb{R}$ as

$$Z(T, i) \coloneqq \frac{\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i)}{\prod w(T)}.$$

If we denote by j_k the first vertex in the ik-walk in $G[T]$, Proposition 4.7 ensures

$$\begin{aligned}
Z(T,i) &= \frac{\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i)}{\prod w(T)} \\
&= \frac{1}{\prod w(T)} \sum_{k \in \Gamma(i)} \mathbb{P}(X_1 = j_k \mid X_0 = i)\mathbb{P}(\mathcal{A}(X) = T - ij_k + ik \mid X_0 = j_k) \\
&= \frac{1}{\prod w(T)} \sum_{k \in \Gamma(i)} \frac{w(ij_k)}{\sum w(\delta(i))} \frac{w(ik)}{w(ik)} \mathbb{P}(\mathcal{A}(X) = T - ij_k + ik \mid X_0 = j_k) \\
&= \sum_{k \in \Gamma(i)} \frac{w(ik)}{\sum w(\delta(i))} \frac{\mathbb{P}(\mathcal{A}(X) = T - ij_k + ik \mid X_0 = j_k)}{\prod w(T - ij_k + ik)} \\
&= \sum_{k \in \Gamma(i)} \frac{w(ik)}{\sum w(\delta(i))} Z(T - ij_k + ik, j_k) \\
&= \sum_{k \in \Gamma(i)} \mathbb{P}(X_1 = k \mid X_0 = i) Z(T - ij_k + ik, j_k).
\end{aligned}$$

Therefore, the function $Z \colon \mathcal{T}_G \times V \to \mathbb{R}$ is harmonic at every vertex of $\mathcal{H}(G)$. Proposition 4.11 ensures that $\mathcal{H}(G)$ is connected, so that Proposition 1.63 implies that Z is constant. In other words, there exists $\alpha \in \mathbb{R}$ such that for every $T \in \mathcal{T}_G$ and for every $i \in V$,

$$\mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i) = \alpha \prod w(T).$$

Therefore,

$$\mathbb{P}(\mathcal{A}(X) = T) = \sum_{i \in V} \mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i)\mathbb{P}(X_0 = i) = \alpha \prod w(T) \sum_{i \in V} \mathbb{P}(X_0 = i) = \alpha \prod w(T).$$

Since G is connected, the Markov chain X is irreducible. Proposition 3.29 ensures that $\mathbb{P}(\mathcal{A}(X) = \perp) = 0$. Proposition 2.37 allow us to conclude that

$$1 = \sum_{T \in \mathcal{T}_G} \mathbb{P}(\mathcal{A}(X) = T) = \sum_{T \in \mathcal{T}_G} \alpha \prod w(T) = \alpha \sum_{T \in \mathcal{T}_G} \prod w(T) = \alpha\, \Phi(G).$$

Thus, $\alpha = \Phi(G)^{-1}$.

Let Y be any random walk on G. Theorem 3.15 ensures that there exists a random walk X on G such that $\mathbb{P}(X_0 = i) = 1/|V|$ for every $i \in V$. We have just proved that $\mathbb{P}\mathcal{A}(X) = TX_0 = i = (\prod w(T))/\,\Phi(G)$. Since X and Y have the same transition matrix, Proposition 4.4 ensures $\{\mathcal{A}(X) = T\}_{\Omega} \in \sigma[\mathrm{Traj}_X]$, so that Proposition 3.19 finishes the proof:

$$\begin{aligned}\mathbb{P}(\mathcal{A}(Y) = T) &= \sum_{i \in V} [\mathbb{P}(Y_0 = i) > 0]\, \mathbb{P}(\mathcal{A}(Y) = T \mid Y_0 = i)\mathbb{P}(Y_0 = i) \\ &= \sum_{i \in V} [\mathbb{P}(Y_0 = i) > 0]\, \mathbb{P}(\mathcal{A}(X) = T \mid X_0 = i)\mathbb{P}(Y_0 = i) \\ &= \frac{\prod w(T)}{\Phi(G)} \sum_{i \in V} [\mathbb{P}(Y_0 = i) > 0]\, \mathbb{P}(Y_0 = i) \\ &= \frac{\prod w(T)}{\Phi(G)}.\end{aligned}$$

□

4.2 Bounds on the Cover Time

The Aldous-Broder algorithm terminates as soon as every vertex has been visited. In other words, the cover time captures the running time of the algorithm. For this reason, we are interested in bounding the expected cover time of a random walk in a graph.

Bounds on the expected cover time of a graph were known before the Aldous-Broder algorithm was developed. Aleliunas et al. [2] gave the first proof of Proposition 4.27 in 1979, the end goal of this section. However, the argument here is different from theirs for two reasons. First, it uses a result first proved by Chandra et al. [15] in 1989 that relates the expected hitting time with the pseudoinverse of the Laplacian of a graph. Second, Proposition 4.25 is a different take on relating the expected cover time with the expected hitting time, which is simpler than the one found in [2] and subsequently used in [15]. An unexpected consequence of such approach is a hint that the cover time can be lower for regular graphs, in Corollary 4.26. The first proof that the expected cover time on regular graphs on n vertices is $O(n^2)$ was given by Kahn et al. in [16], and Feige [8] lowered the constant.

The main idea in this section is to "transport" results about stationary random walks into arbitrary ones. It is not hard to imagine that some properties of random walks are simpler to study when one is dealing with the stationary distribution. It is more interesting to understand how to exploit this special case to prove results about random walks with no hypothesis on the initial distribution. We will do such a thing with the next results, up until Corollary 4.19.

Proposition 4.13. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, and let Y be a stationary walk on G. For every $E \in \sigma(\mathcal{O}_V)$,

$$\mathbb{P}(\mathrm{Traj}_X \in E \mid X_0 = i) = \mathbb{P}(\mathrm{Traj}_Y \in E \mid Y_0 = i),$$

whenever $i \in V$ is such that both conditional probabilities are defined.

Proof. Since X and Y are random walks on the same graph, by definition both have the same transition probabilities. It is enough to apply Proposition 3.19. □

Proposition 4.14. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a stationary walk on G. Denote by $\pi \in \mathbb{R}^V_{++}$ the stationary distribution of G. For every $i \in V$,

$$\mathbb{E}[\mathrm{Arr}_{X,i}] = 1/\pi_i.$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space in which X is defined. Since X is stationary, for every $t \in \mathbb{N}$,

$$\mathbb{P}(\mathrm{Arr}_{X,i} = t) = \mathbb{P}(X_t = i) \prod_{k=0}^{t-1} \mathbb{P}(X_k \neq i) = \mathbb{P}(X_t = i) \prod_{k=0}^{t-1} (1 - \mathbb{P}(X_k = i)) = \pi_i (1 - \pi_i)^{t-1}.$$

Therefore

$$\begin{aligned} \mathbb{E}[\mathrm{Arr}_{X,i}] &= \sum_{t \in \mathbb{N}} t\, \mathbb{P}(\mathrm{Arr}_{X,i} = t) \\ &= \sum_{t \in \mathbb{N}} t\, \pi_i (1 - \pi_i)^{t-1} \\ &= \pi_i \sum_{t \in \mathbb{N}} t\, (1 - \pi_i)^{t-1} && \text{by Proposition 1.67} \\ &= \pi_i \frac{1}{\pi_i^2} = \frac{1}{\pi_i} && \text{by Theorem 1.70.} \end{aligned}$$

□

Proposition 4.15 (Law of Total Expectation)**.** Let X be random variable in $(\Omega, \mathcal{F}, \mathbb{P})$, taking values with nonzero probabilities only on $\mathbb{N}$. Let $\mathcal{E} \subseteq \mathcal{F}$ be a partition of Ω in events, with $|\mathcal{E}|$ finite. Then

$$\mathbb{E}[X] = \sum_{E \in \mathcal{E}} [\mathbb{P}(E) > 0]\, \mathbb{P}(E) \mathbb{E}\left[X | E\right].$$

Proof. Expand the summation, condition the probabilities on the events of the statement, and manipulate the series accordingly:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{t \in \mathbb{N}} t\, \mathbb{P}(X = t) \\ &= \sum_{t \in \mathbb{N}} t \left(\sum_{E \in \mathcal{E}} \mathbb{P}(X = t, E) \right) \\ &= \sum_{t \in \mathbb{N}} t \left(\sum_{E \in \mathcal{E}} [\mathbb{P}(E) > 0]\, \mathbb{P}(X = t \mid E) \mathbb{P}(E) \right) \\ &= \sum_{E \in \mathcal{E}} \sum_{t \in \mathbb{N}} [\mathbb{P}(E) > 0]\, t\, \mathbb{P}(X = t \mid E) \mathbb{P}(E) && \text{by Theorem 1.68} \\ &= \sum_{E \in \mathcal{E}} [\mathbb{P}(E) > 0]\, \mathbb{P}(E) \sum_{t \in \mathbb{N}} t\, \mathbb{P}(X = t \mid E) && \text{by Proposition 1.67} \\ &= \sum_{E \in \mathcal{E}} [\mathbb{P}(E) > 0]\, \mathbb{P}(E) \mathbb{E}\left[X | E\right]. \end{aligned}$$

□

Corollary 4.16. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G. Let $i \in V$ be such that $\mathbb{P}(X_0 = i) > 0$. For every $j \in V$,

$$\mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = i\right] < \infty.$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space in which X is defined. Let Y be a stationary walk on G, in a probability space $(\Omega', \mathcal{G}, \mathbb{P}_Y)$. Proposition 4.13 ensures that for every $t \in \mathbb{N}$,

$$\mathbb{P}_Y(\mathrm{Arr}_{Y,j} = t \mid Y_0 = i) = \mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = i).$$

Therefore

$$\mathbb{E}\left[\mathrm{Arr}_{Y,j}|Y_0=i\right]=\sum_{t\in\mathbb{N}}t\,\mathbb{P}_Y(\mathrm{Arr}_{Y,j}=t\mid Y_0=i)=\sum_{t\in\mathbb{N}}t\,\mathbb{P}(\mathrm{Arr}_{X,j}=t\mid X_0=i)=\mathbb{E}\left[\mathrm{Arr}_{X,j}|X_0=i\right].$$

Since G is connected, Proposition 3.29 ensures that for every $i\in V$, the random variables $\mathrm{Arr}_{Y,i}$ and $\mathrm{Arr}_{X,i}$ take only values in $\mathbb{N}$ with nonzero probability. Therefore, Proposition 4.15 applied to the events $\{Y_0=i\}_{\Omega'}$ ensures that

$$\mathbb{E}[\mathrm{Arr}_{Y,j}]=\sum_{i\in V}\mathbb{P}(Y_0=i)\mathbb{E}\left[\mathrm{Arr}_{Y,j}|Y_0=i\right]\geq\pi_i\mathbb{E}\left[\mathrm{Arr}_{Y,j}|Y_0=i\right]=\pi_i\,\mathbb{E}\left[\mathrm{Arr}_{X,j}|X_0=i\right].$$

Since $\pi_i>0$, the fact that $\mathbb{E}\left[\mathrm{Arr}_{X,j}|X_0=i\right]=\infty$, would then imply that $\mathbb{E}[\mathrm{Arr}_{Y,j}]=\infty$, contrary to Propositon 4.14. Hence, $\mathbb{E}\left[\mathrm{Arr}_{X,j}|X_0=i\right]$ is finite. □

The corollary just proved is important, since it ensures that when working with arrival times, we are actually working with real numbers. It motivates the following definition.

Definition 4.17. Let $G=(V,E,w)$ be a simple, connected and weighted graph, with $w\in\mathbb{R}^E_{++}$. Let X be a stationary walk on G. The *hitting time matrix of* G, denoted $H_G\in\mathbb{R}^{V\times V}$, is defined as

$$(H_G)_{ij}\coloneqq\mathbb{E}[\mathrm{Arr}_{X,j}\mid X_0=i],$$

for every $i,j\in V$.

Let $G=(V,E,w)$ be a simple, connected and weighted graph, with $w\in\mathbb{R}^E_{++}$. Let X be a random walk on G. Proposition 4.13 ensures that, whenever $i\in V$ is such that the conditional probability is defined,

$$\mathbb{E}\left[\mathrm{Arr}_{X,j}|X_0=i\right]=(H_G)_{ij},$$

for every $j\in V$. Hence, the hitting time matrix of G captures the expected hitting time of any random walk on G.

Corollary 4.18. Let $G=(V,E,w)$ be a simple, connected and weighted graph, with $w\in\mathbb{R}^E_{++}$. Let $\pi\in\mathbb{R}^V_{++}$ be the stationary distribution of G. Then

$$H_G^{\mathsf{T}}\pi=2w^{\mathsf{T}}\mathbb{1}D_G^{-1}\mathbb{1}.$$

Proof. Let X be a stationary walk on G, in a probability space $(\Omega,\mathcal{F},\mathbb{P})$. For every $i\in V$, Proposition 4.13, Proposition 4.14, and Proposition 3.9 ensure that

$$e_i^{\mathsf{T}}H_G^{\mathsf{T}}\pi=\sum_{j\in V}\mathbb{E}[\mathrm{Arr}_{X,i}\mid X_0=j]\,\pi_j=\mathbb{E}[\mathrm{Arr}_{X,i}]=1/\pi_i=e_i^{\mathsf{T}}\big(2w^{\mathsf{T}}\mathbb{1}D_G^{-1}\mathbb{1}\big).$$

Hence, $H_G^{\mathsf{T}}\pi=2w^{\mathsf{T}}\mathbb{1}D_G^{-1}\mathbb{1}$. □

Corollary 4.19. Let $G=(V,E)$ be a simple, connected and k-regular graph. Then

$$H_G^{\mathsf{T}}\mathbb{1}=|V|^2\,\mathbb{1}.$$

Proof. Let $\pi\in\mathbb{R}^V_{++}$ be the stationary distribution of G. Since G is k-regular, we have that $\mathbb{1}=|V|\,\pi$, that $2\,|E|=k\,|V|$, and that $D_G=kI$. Corollary 4.18, applied with $w=\mathbb{1}$, then finishes the proof:

$$H_G\mathbb{1}=|V|\,H_G^{\mathsf{T}}\left(\frac{1}{|V|}\mathbb{1}\right)=|V|\,H_G^{\mathsf{T}}\pi=|V|\,\big(2\mathbb{1}^{\mathsf{T}}\mathbb{1}D_G^{-1}\mathbb{1}\big)=|V|\,\big(2\,|E|\,D_G^{-1}\mathbb{1}\big)=|V|\left(\frac{k\,|V|}{k}\mathbb{1}\right)=|V|^2\,\mathbb{1}.\quad\square$$

Proposition 4.20. Let $G=(V,E,w)$ be a simple, connected and weighted graph, with $w\in\mathbb{R}^E_{++}$. Let X be a random walk on G in a probability space $(\Omega,\mathcal{F},\mathbb{P})$. Then for every $k,j\in V$,

$$\sum_{t\in\mathbb{N}}t\,\mathbb{P}(\mathrm{Arr}_{X,j}=t-1\mid X_0=k)=\sum_{t\in\mathbb{N}}(1+t)\mathbb{P}(\mathrm{Arr}_{X,j}=t\mid X_0=k).$$

Proof. Define $(\alpha_t)_{t\in\mathbb{N}}$ and $(\beta_t)_{t\in\mathbb{N}}$ as

$$\begin{aligned}\alpha_t &:= t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k)\\ \beta_t &:= (1+t)\mathbb{P}(\mathrm{Arr}_{X,j} = t, \mid X_0 = k).\end{aligned}$$

Note that, for every $T \in \mathbb{N}$,

$$\sum_{t=0}^{T} \alpha_t = \sum_{t=0}^{T} t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k) = \sum_{t=0}^{T-1}(1+t)\mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = k) = \sum_{t=0}^{T-1} \beta_k.$$

Therefore,

$$\left\{\sum_{t=0}^{T} \alpha_t : T \in \mathbb{N}\right\} = \left\{\sum_{t=0}^{T} \beta_t : T \in \mathbb{N}\right\}.$$

Equation (1.65) then finishes the proof. □

Proposition 4.21. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a stationary walk on G. For every $i \in V$ and for every $j \in V \setminus \{i\}$,

$$\mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = i\right] = 1 + \sum_{k\in V} \mathbb{P}(X_1 = k \mid X_0 = i)\mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = k\right].$$

Proof. Let $P \in \mathbb{R}^{V\times V}$ be the transition matrix of X. For every $t \in \mathbb{N}$, Proposition 3.21 ensures that $\{\mathrm{Arr}_{X,j} = t\}_\Omega \in \sigma[\mathrm{Traj}_X]$. Moreover, Proposition 3.27 implies

$$\begin{aligned}\{\mathrm{Arr}_{X,j} = t, X_0 = i\}_\Omega &= \bigcup_{k\in V} \{\mathrm{Arr}_{X,j} = t, Y_0 = k, X_0 = i\}_\Omega\\ &= \bigcup_{k\in V} \{\mathrm{Arr}_{Y,j} = t-1, Y_0 = k, X_0 = i\}_\Omega.\end{aligned}$$

Furthermore, note that Proposition 3.19 and Proposition 3.28 ensure that for every $k \in V$,

$$\begin{aligned}\mathbb{P}(\mathrm{Arr}_{X,j} = t, Y_0 = k \mid X_0 = i) &= [P_{ik} > 0]\, P_{ik}\, \mathbb{P}(\mathrm{Arr}_{Y,j} = t \mid Y_0 = k, X_0 = i)\\ &= [P_{ik} > 0]\, P_{ik}\, \mathbb{P}(\mathrm{Arr}_{Y,j} = t \mid Y_0 = k)\\ &= [P_{ik} > 0]\, P_{ik}\, \mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k)\\ &= P_{ik}\, \mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k).\end{aligned}$$

In the last equality we used the fact that X is stationary, which implies that $\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k)$ is always defined. Therefore,

$$\begin{aligned}\mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = i\right] &= \sum_{t\in\mathbb{N}} t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = k) = \sum_{t\in\mathbb{N}} t\left(\sum_{k\in V} \mathbb{P}(\mathrm{Arr}_{X,j} = t, Y_0 = k \mid X_0 = i)\right)\\ &= \sum_{t\in\mathbb{N}} t\left(\sum_{k\in V} P_{ik}\,\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k)\right)\\ &= \sum_{k\in V}\sum_{t\in\mathbb{N}} P_{ik}\, t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k) && \text{by Theorem 1.68}\\ &= \sum_{k\in V} P_{ik} \sum_{t\in\mathbb{N}} t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t-1 \mid X_0 = k) && \text{by Proposition 1.67}\\ &= \sum_{k\in V} P_{ik} \sum_{t\in\mathbb{N}} (1+t)\mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = k) && \text{by Proposition 4.20}\\ &= \sum_{k\in V} P_{ik}\left(\sum_{t\in\mathbb{N}} \mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = i) + \sum_{t\in\mathbb{N}} t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = i)\right) && \text{by Theorem 1.68}\\ &= \sum_{k\in V} P_{ik}(1 + \mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = k\right]) = 1 + \sum_{k\in V} P_{ik}\mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = k\right].\end{aligned}$$

□

Proposition 4.22. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let P be the transition matrix of G. Then

$$(I - P)H_G = \mathbb{1}\mathbb{1}^\mathsf{T} - 2w^\mathsf{T}\mathbb{1}D_G^{-1}.$$

Proof. Define $A \in \mathbb{R}^{V \times V}$ as $A \coloneqq (I - P)H_G - \mathbb{1}\mathbb{1}^\mathsf{T}$. Suffices then to show that $A = -2w^\mathsf{T}\mathbb{1}D_G^{-1}$. Proposition 4.21 ensures that for every $i, j \in V$,

$$(H_G)_{ij} = 1 + \sum_{k \in V} P_{ik}(H_G)_{kj}.$$

Hence, A is diagonal. Let $\pi \in \mathbb{R}^V_{++}$ be the stationary distribution of G. Since $\pi = P^\mathsf{T}\pi$, we have that $\pi \in \operatorname{Null}((I - P)^\mathsf{T})$. Therefore

$$\begin{aligned} A^\mathsf{T}\pi &= ((I - P)H_G - \mathbb{1}\mathbb{1}^\mathsf{T})^\mathsf{T}\pi \\ &= H_G^\mathsf{T}(I - P)^\mathsf{T}\pi - \mathbb{1}\mathbb{1}^\mathsf{T}\pi \\ &= -\mathbb{1}. \end{aligned}$$

Since A is diagonal, we conclude that for every $i \in V$,

$$A_{ii} = -1/\pi_i = -\big(2w^\mathsf{T}\mathbb{1}D_G^{-1}\big)_{ii}.$$

Therefore, $A = -2w^\mathsf{T}\mathbb{1}D_G^{-1}$. □

The next two results were first proved in [15] with a slightly different language. For every graph, the authors define an electric circuit related to it, and reason about the voltages and currents in it. Here, every argument about voltage and current of such circuits is made via algebraic manipulations of the Laplacian and its pseudoinverse. Doyle and Snells [7] and Wagner [20] have interesting texts on the relation between electric circuits and the Laplacian of a graph.

Proposition 4.23. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Then for every $i, j \in V$,

$$(H_G)_{ij} = (e_i - e_j)^\mathsf{T} L_G^\dagger\big(D_G\mathbb{1} - 2w^\mathsf{T}\mathbb{1}e_j\big).$$

Proof. Let $P \in \mathbb{R}^{V \times V}$ be the transition matrix of G. Proposition 1.57 ensures that

$$L_G = D_G - A_G = D_G(I - D_G^{-1}A) = D_G(I - P).$$

Moreover, Proposition 4.22 ensures that

$$(I - P)H_G = \mathbb{1}\mathbb{1}^\mathsf{T} - 2w^\mathsf{T}\mathbb{1}D_G^{-1}.$$

Left multiply both sides by $L_G^\dagger D_G$ to conclude that

$$L_G^\dagger L_G H_G = L_G^\dagger\big(D_G\mathbb{1}\mathbb{1}^\mathsf{T} - 2w^\mathsf{T}\mathbb{1}\big).$$

Proposition 1.58 ensures that $L_G^\dagger L_G = P_{\operatorname{span}(\mathbb{1})^\perp}$. Since $e_i - e_j \in \operatorname{span}(\mathbb{1})^\perp$, and $P_{\operatorname{span}(\mathbb{1})^\perp}$ is orthogonal,

$$(e_i - e_j)^\mathsf{T} P_{\operatorname{span}(\mathbb{1})^\perp} = \Big(P^\mathsf{T}_{\operatorname{span}(\mathbb{1})^\perp}(e_i - e_j)\Big)^\mathsf{T} = \big(P_{\operatorname{span}(\mathbb{1})^\perp}(e_i - e_j)\big)^\mathsf{T} = (e_i - e_j)^\mathsf{T}.$$

Hence,

$$(e_i - e_j)^\mathsf{T} P_{\operatorname{span}(\mathbb{1})^\perp} H_G e_j = (e_i - e_j)^\mathsf{T} H_G e_j = (e_i - e_j)^\mathsf{T} L_G^\dagger(D_G\mathbb{1} - 2w^\mathsf{T}\mathbb{1}e_j).$$

However, since $(H_G)_{jj} = 0$,

$$(e_i - e_j)^\mathsf{T} P_{\operatorname{span}(\mathbb{1})^\perp} H_G e_j = (e_i - e_j)^\mathsf{T} H_G e_j = (H_G)_{ij} - (H_G)_{jj} = (H_G)_{ij}.$$ □

The next proposition relates the entries in the hitting time matrix with a more familiar quantity.

Theorem 4.24. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Then for every $i, j \in V$,

$$(H_G)_{ij} + (H_G)_{ji} = 2w^\mathsf{T}\mathbb{1}(e_j - e_i)^\mathsf{T} L_G^\dagger(e_j - e_i).$$

Proof. Proposition 4.23 ensures that

$$(H_G)_{ji} = (e_j - e_i)^\mathsf{T} L_G^\dagger(d - 2w^\mathsf{T}\mathbb{1}e_i) = (e_i - e_j)^\mathsf{T} L_G^\dagger(2w^\mathsf{T}\mathbb{1}e_i - d).$$

Therefore,

$$\begin{aligned}(H_G)_{ij} + (H_G)_{ji} &= (e_i - e_j)^\mathsf{T} L_G^\dagger(d - 2w^\mathsf{T}\mathbb{1}e_j)\\ &\quad + (e_i - e_j)^\mathsf{T} L_G^\dagger(2w^\mathsf{T}\mathbb{1}e_i - d)\\ &= (e_i - e_j)^\mathsf{T} L_G^\dagger(2w^\mathsf{T}\mathbb{1}e_i - 2w^\mathsf{T}\mathbb{1}e_j)\\ &= 2w^\mathsf{T}\mathbb{1}(e_i - e_j)^\mathsf{T} L_G^\dagger(e_i - e_j).\end{aligned}$$

□

Proposition 4.25. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $i \in V$ be such that $\mathbb{P}(X_0 = i) > 0$. Then

$$\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq e_i^\mathsf{T} H_G \mathbb{1}.$$

Proof. Define

$$S := \{\, j \in V : \mathbb{P}(\mathrm{Last}_X = j, X_0 = i) > 0\}.$$

Note that $\{X_0 = i\}_\Omega = \bigcup_{j \in S}\{\mathrm{Last}_X = j, X_0 = i\}_\Omega$. In other words, the events $\{\mathrm{Last}_X = j, X_0 = i\}_\Omega$, for $j \in S$, are a partition in events of $\{X_0 = i\}_\Omega$. Hence, Proposition 4.15 implies

$$\begin{aligned}\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] &= \sum_{j \in S} \mathbb{P}(\mathrm{Last}_X = j \mid X_0 = i)\mathbb{E}\left[\mathrm{Cov}_X | \mathrm{Last}_X = j, X_0 = i\right]\\ &= \sum_{j \in S} \mathbb{P}(\mathrm{Last}_X = j \mid X_0 = i)\left(\sum_{t \in \mathbb{N}} t\,\mathbb{P}(\mathrm{Cov}_X = t \mid \mathrm{Last}_X = j, X_0 = i)\right)\\ &= \sum_{j \in S}\sum_{t \in \mathbb{N}} t\,\mathbb{P}(\mathrm{Cov}_X = t \mid \mathrm{Last}_X = j, X_0 = i)\mathbb{P}(\mathrm{Last}_X = j \mid X_0 = i) \qquad \text{by Proposition 1.67}\\ &= \sum_{j \in S}\sum_{t \in \mathbb{N}} t\,\mathbb{P}(\mathrm{Cov}_X = t, \mathrm{Last}_X = j \mid X_0 = i)\end{aligned}$$

Moreover, for every $j \in S$ and $t \in \mathbb{N}$, we have that

$$\{\mathrm{Cov}_X = t, \mathrm{Last}_X = j, X_0 = i\}_\Omega = \{\mathrm{Arr}_{X,j} = t, \mathrm{Last}_X = j, X_0 = i\}_\Omega \subseteq \{\mathrm{Arr}_{X,j} = t, X_0 = i\}_\Omega,$$

Hence, $\mathbb{P}(\mathrm{Cov}_X = t, \mathrm{Last}_X = j \mid X_0 = i) \leq \mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = i)$. Therefore,

$$\begin{aligned}\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] &= \sum_{j \in S}\sum_{t \in \mathbb{N}} t\,\mathbb{P}(\mathrm{Cov}_X = t, \mathrm{Last}_X = j \mid X_0 = i)\\ &\leq \sum_{j \in S}\sum_{t \in \mathbb{N}} t\,\mathbb{P}(\mathrm{Arr}_{X,j} = t \mid X_0 = i) \qquad \text{by Proposition 1.66}\\ &\leq \sum_{j \in S} \mathbb{E}\left[\mathrm{Arr}_{X,j} | X_0 = i\right] \leq e_i^\mathsf{T} H_G \mathbb{1}.\end{aligned}$$

□

Corollary 4.26. Let $G = (V, E)$ be a connected, k-regular graph, and let X be a random walk on G. There exists $i \in V$ such that

$$\mathbb{E}[\mathrm{Cov}_X \mid X_0 = i] \leq |V|^2.$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space in which X is defined. Proposition 4.25 and Corollary 4.19 ensure

$$\sum_{i\in V} \mathbb{E}[\mathrm{Cov}_X \mid X_0 = i] \leq \sum_{i \in V} e_i^\mathsf{T} H_G \mathbb{1} = \left(\mathbb{1}^\mathsf{T} H_G\right)\mathbb{1} = \left(|V|^2 \, \mathbb{1}\right)^\mathsf{T} \mathbb{1} = |V|^3 .$$

In other words,

$$\sum_{i\in V} \left(\frac{1}{|V|} \mathbb{E}[\mathrm{Cov}_X \mid X_0 = i]\right) \leq |V|^2 .$$

Hence, Proposition 1.4 finishes the proof. □

Proposition 4.27. Let $G = (V, E, w)$ be a simple, connected and weighted graph, with $w \in \mathbb{R}^E_{++}$. Let X be a random walk on G. Denote by $W \coloneqq \min_{ij\in E} w(ij)$. For every $i \in V$ such that $\mathbb{P}(X_0 = i) > 0$,

$$\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq \frac{2w^\mathsf{T}\mathbb{1}}{W}(|V| - 1).$$

Proof. Let $j \in V$ be such that $(H_G)_{ij}$ is maximal. Then Proposition 4.25 and the fact that $(H_G)_{ii} = 0$ ensure

$$\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq e_i^\mathsf{T} H_G \mathbb{1} = \sum_{k\in V\setminus\{i\}} (H_G)_{ik} \leq \sum_{k\in V\setminus\{i\}} (H_G)_{ij} = (|V| - 1)(H_G)_{ij}.$$

Since $(H_G)_{ji} \geq 0$, Theorem 4.24 implies

$$\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq (|V| - 1)((H_G)_{ij} + (H_G)_{ji}) = (|V| - 1)2w^\mathsf{T}\mathbb{1}(e_i - e_j)^\mathsf{T} L_G^\dagger (e_i - e_j).$$

If $ij \notin E$, define the graph $G' \coloneqq (V, E \cup \{ij\}, w')$, with w' equal to w in every edge in E, and with $w'(ij) \coloneqq W$. If $ij \in E$, define $G' \coloneqq G$. Note that $\mathcal{A}(G')$, as defined in Section 2.4, is a random variable as required in the statement of Theorem 2.50, and that this theorem implies that

$$\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq (|V| - 1)2w^\mathsf{T}\mathbb{1}\frac{\mathbb{P}(ij \in \mathcal{A}(G'))}{w'(ij)} \leq \frac{2w^\mathsf{T}\mathbb{1}}{W}(|V| - 1).$$ □

Corollary 4.28. Let $G = (V, E)$ be a simple graph. For every random walk X on G,

$$\mathbb{E}[\mathrm{Cov}_X] \leq 2\,|E|\,(|V| - 1).$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space in which X is defined. Define

$$S \coloneqq \{\, i \in V : \mathbb{P}(X_0 = i) > 0\}.$$

We have that $\sum_{i\in S} \mathbb{P}(X_0 = i) = 1$, and that the events $\{X_0 = i\}_\Omega$, for every $i \in S$, are a partition Ω. Hence, Proposition 4.15 and Proposition 4.27 imply

$$\mathbb{E}[\mathrm{Cov}_X] = \sum_{i\in S} \mathbb{P}(X_0 = i)\mathbb{E}\left[\mathrm{Cov}_X | X_0 = i\right] \leq \sum_{i\in S} \mathbb{P}(X_0 = i) 2\,|E|\,(|V| - 1) = 2\,|E|\,(|V| - 1).$$ □

References

[1] D. J. Aldous. “The random walk construction of uniform spanning trees and uniform labelled trees”. In: *SIAM J. Discrete Math.* 3.4 (1990), pages 450–465. URL: http://dx.doi.org/10.1137/0403039 (cited on pages 1, 62).

[2] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, and C. Rackoff. “Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems”. In: (Nov. 1979), pages 218–223 (cited on page 67).

[3] A. Asadpour, M. X. Goemans, A. Mądry, S. Oveis Gharan, and A. Saberi. “An $O(\log n/\log\log n)$-approximation algorithm for the asymmetric traveling salesman problem”. In: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms.* SIAM, Philadelphia, PA, 2010, pages 379–389 (cited on page 1).

[4] N. Bourbaki. *Algebra I.* Volume Chapters 1-3. Springer, 1998 (cited on page 29).

[5] A. Broder. “Generating random spanning trees”. In: *Proceedings of the 30th Annual Symposium on Foundations of Computer Science, FOCS 1989.* 1989, pages 442–447 (cited on pages 1, 62).

[6] K. Conrad. *The Sign of a Permutation.* URL: http://www.math.uconn.edu/~kconrad/blurbs/grouptheory/sign.pdf (cited on page 8).

[7] P. G. Doyle and J. L. Snell. “Random walks on electric networks”. In: (2009). URL: https://math.dartmouth.edu/~doyle/docs/walks/walks.pdf (cited on page 71).

[8] U. Feige. *Collecting Coupons on Trees, and the Analysis of Random Walks.* Technical report. of the Weizmann Institute, 1994 (cited on page 67).

[9] A. Frieze, N. Goyal, L. Rademacher, and S. Vempala. “Expanders via random spanning trees”. In: *SIAM J. Comput.* 43.2 (2014), pages 497–513. URL: http://dx.doi.org/10.1137/120890971 (cited on page 1).

[10] O. Häggström. *Finite Markov Chains and Algorithmic Applications.* Cambridge University Press, 2002 (cited on page 53).

[11] P. Halmos. *Finite-Dimensional Vector Spaces.* Undergraduate Texts in Mathematics. Springer New York, 2012. URL: https://books.google.com.br/books?id=yzr0BwAAQBAJ (cited on page 8).

[12] P. Halmos. *Measure Theory.* Graduate Texts in Mathematics. Springer New York, 1976. URL: https://books.google.com.br/books?id=-Rz7q4jikxUC (cited on page 57).

[13] S. Hoory, N. Linial, and A. Wigderson. “Expander graphs and their applications”. In: *Bull. Amer. Math. Soc. (N.S.)* 43.4 (2006), 439–561 (electronic). URL: http://dx.doi.org/10.1090/S0273-0979-06-01126-8 (cited on page 1).

[14] N. Jacobson. *Basic Algebra.* Basic Algebra v. 1. W.H. Freeman, 1985. URL: https://books.google.com.br/books?id=-G0PAQAAMAAJ (cited on page 8).

[15] A. K. Chandra, P. Raghavan, W. Ruzzo, R. Smolensky, and P. Tiwari. “The Electrical Resistance of a Graph Captures its Commute and Cover Times (Detailed Abstract)”. In: (Jan. 1989), pages 574–586 (cited on pages 67, 71).

[16] J. D. Kahn, N. Linial, N. Nisan, and M. E. Saks. “The Electrical Resistance of a Graph Captures its Commute and Cover Times (Detailed Abstract)”. In: *Journal of Theoretical Probability* 2.1 (1989), pages 121–128 (cited on page 67).

[17] G. Kirchhoff. “Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird”. In: *Ann. Phys. und Chem.* 72 (1847), pages 497–508 (cited on pages 1, 29).

[18] V. G. Kulkarni. “Generating Random Combinatorial Objects”. In: *J. Algorithms* 11.2 (May 1990), pages 185–207. URL: http://dx.doi.org/10.1016/0196-6774(90)90002-V (cited on pages 1, 29, 48).

[19] T. Tao. *An Introduction to Measure Theory*. Volume vol. 126. Graduate Studies in Mathematics. American Mathematical Society, 2011 (cited on page 21).

[20] D. G. Wagner. “Combinatorics of Electrical Networks”. In: (2009). URL: http://www.math.uwaterloo.ca/~dgwagner/Networks.pdf (cited on page 71).