Contents

1 Introduction **2**

2 Preliminaries **3**
2.1 Order . . . 3
2.2 Graph theory . . . 3
2.3 Linear algebra . . . 6
2.3.1 Basic definitions . . . 6
2.3.2 Projections and direct sum . . . 7
2.4 The Laplacian . . . 8
2.5 Harmonic Functions . . . 9
2.6 Measure Theory . . . 10

I Laplacian-Based Algorithms **12**

3 The Naive Algorithm **13**
3.1 Cauchy-Binet formula . . . 13
3.2 Matrix Tree Theorems . . . 19
3.3 The Algorithm . . . 22

4 The Harvey-Xu Algorithm **27**
4.1 The Moore-Penrose pseudoinverse . . . 27
4.2 Effective Resistances as Marginal Probabilities . . . 31
4.3 The Harvey and Xu algorithm . . . 32

II Random Walk-Based Algorithms **33**

5 Random Walks **34**
5.1 Markov Chains . . . 34
5.2 Arrival time and cover time . . . 34

6 The Aldous-Broder Algorithm **36**

Chapter 1

Introduction

Seriously, the stuff here is really cool and everybody should pay me to study it.

Chapter 2

Preliminaries

2.1 Order

Definition 2.1. Given a set A, a relation $\leq$ in A is a *partial order* in it if

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

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

(iii) the relation is *transitive*, that is, 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 2.2. A partially ordered set $(A, \leq)$ is said to be *totally ordered* if the relation is *total*, that is, if for every $a, b \in A$ it holds that $a \leq b$ or $b \leq a$.

2.2 Graph theory

Definition 2.3. A *graph* is an ordered triple $G = (V, E, \psi)$, 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* on 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 $\{i, j\} \in \psi(E)$.

Definition 2.4. A *digraph* is an ordered triple $D = (V, A, \psi)$, where V and E 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 the difference in their 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 $\psi(a) = (i, j)$, i is said to be the *tail* of the arc, j is said to be the *head* of the arc, and a is said to be incident on i and to be incident on j.

There's now a need to stop and pay respect to tradition. There are some notations that are widespread for its 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's hoped the reader will notice and correctly parse it as (i, j) when it's an arc, and as $\{i, j\}$ when it's an edge.

Also, graphs are commonly thought of as a symmetric relation E on a finite set V. This interpretation usually cast 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, and only when this happens, the incidence function will be omitted, and the graph will 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.

Definition 2.5. 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 $(i, j) \mapsto \{i, j\}$. The *underlying graph* of D is the graph $G \coloneqq (V, E, \phi)$, where $E = A$ and $\phi = \pi \circ \psi$.

The function π in the above definition encrypts 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.

Definition 2.6. An *weighted graph* is a pair (G, w), where $G = (V, E, \psi)$ is a graph and $w\colon E \to \mathbb{R}$. It can also be denoted as $G = (V, E, \psi, w)$.

Definition 2.7. An *weighted digraph* is a pair (D, w), where $D = (V, A, \psi)$ is a digraph and $w\colon A \to \mathbb{R}$. It can also be denoted as $D = (V, A, \psi, w)$.

Given an weigthed digraph (D, w), the weigthed graph (G, w), with G being the underlying graph of D, will be called *underlying weighted graph* of (D, w).

Definition 2.8. Let G be a graph. Two distinct vertices i and j are said to be *connected* if either they are adjacent, or there exists a third vertex k adjacent to i that is connected to j. A graph is said to be connected if every pair of distinct vertices is connected.

Definition 2.9. 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.

Definition 2.10. 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", is a complete 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.

Definition 2.11. A *component* of a graph G is a maximal connected subgraph.

Definition 2.12. A *walk* on a graph G is a finite alternating sequence of vertices and edges $(u_0, e_0, u_1, \ldots, e_{l-1}, u_l)$ such that, for every $0 \le i < l$,

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

The integer l is called the *length* of the walk, and is precisely the number of edges in it.

Similarly, a walk on a digraph D is a finite alternating sequence of vertices and arcs $(u_0, a_0, u_1, \ldots, a_{l-1}, u_l)$ such that, for every $0 \le i < l$,

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

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

Definition 2.13. A *trail* on a graph is a walk $(u_0, e_0, \ldots, e_{l-1}, u_l)$ on it such that the map $i \mapsto e_i$, defined on $\{0, \ldots, l-1\}$, is injective.

Definition 2.14. A *trail* on a digraph is a walk $(u_0, a_0, \ldots, a_{l-1}, u_l)$ on it such that the map $i \mapsto a_i$, defined on $\{0, \ldots, l-1\}$, is injective.

Definition 2.15. A *cycle* on a graph is a trail $(u_0, e_0, \ldots, e_{l-1}, u_l)$ on it such that $u_0 = u_l$.

Definition 2.16. A *path* on a graph is a walk $(u_0, e_0, \ldots, e_{l-1}, u_l)$ on it such that the map $i \mapsto u_i$, defined on $\{0, \ldots, l\}$ is injective.

Definition 2.17. A *path* on a digraph is a walk $(u_0, a_0, \ldots, a_{l-1}, u_l)$ on it such that the map $i \mapsto u_i$, defined on $\{0, \ldots, l\}$, is injective.

Definition 2.18. A *spanning subgraph* $H = (S, F, \psi)$ of $G = (V, E, \psi)$ is a subgraph such that $S = V$.

Definition 2.19. Let $G = (V, E, \psi)$ be a graph. Let $S \subseteq V$. The *spanning subgraph generated by* S is the subgraph $G[S] = (V, F)$, where

$$F := \{ e \in E : \phi(e) \subseteq S \}.$$

Definition 2.20. A *spanning tree* of a graph G 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 $\mathcal{T}_G$.

Definition 2.21. A *tree* is a graph $G = (V, E, \psi)$ such that (V, E) itself is a spanning tree.

Definition 2.22. Let $G = (V, E, \psi)$ be a graph. The *neighborhood function* is the function $\delta \colon V \to 2^E$ defined by $i \mapsto \{ e \in E : i \in \psi(e) \}$. The integer $|\delta(i)|$ is called the *degree* of the vertex i.

Definition 2.23. Let $D = (V, A, \psi)$ be a digraph. The *in-neighborhood function* is the function $\delta^{\text{in}} \colon V \to 2^A$ defined by $i \mapsto \{ e \in E : \exists j \in V \quad \psi(e) = ji \}$. The integer $\left|\delta^{\text{in}}(i)\right|$ is called the *in-degree* of vertex i.

Definition 2.24. Let $D = (V, A, \psi)$ be a digraph. The *out-neighborhood function* is the function $\delta^{\text{out}} \colon V \to 2^A$ defined by $i \mapsto \{ e \in E : \exists j \in V \quad \psi(e) = ij \}$. The integer $|\delta^{\text{out}}(i)|$ is called the *out-degree* of vertex i.

Definition 2.25. A *s-arborescence* is a digraph $D = (V, A, \psi)$ such that its underlying graph G is a tree and such that, for every $i \in V$,

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

The collection of sets of edges B such that (V, B) is a s-arborescence is denoted as $\mathcal{T}_D(s)$.

The following theorem will be used tacitly in inductions involving trees.

Theorem 2.26. Let $T = (V, E)$ be a tree with $|V| > 2$. Then there are at least 2 vertices of T with degreee 1.

Proof. First, note that for every $f \in \mathbb{R}^V$, if there is a real number $\alpha \in \mathbb{R}$ such that $\alpha < f(i)$ for every $i \in V$, then $\alpha < 1/|V| \sum_{i \in V} f(i)$. This is clear from

$$|V| \alpha = \sum_{i \in V} \alpha < \sum_{i \in V} f(i).$$

The contrapositive of this statement is that, for every $f \in \mathbb{R}^V$, there is a $k \in V$ such that

$$f(k) \leq \frac{1}{|V|} \sum_{i \in V} f(i).$$

This 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 the 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. 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. □

Vertices in a tree whose degree equals to 1 are called *leafs*.

Theorem 2.27. Let $D = (V, A)$ be an i-arborescence with $|V| > 2$. Then there is $j \in V \setminus \{i\}$ 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 $\delta^{out}(j) = \varnothing$. □

2.3 Linear algebra

2.3.1 Basic definitions

Definition 2.28. An *Euclidean space* $\mathbb{E}$ is a pair $(V, \langle\cdot,\cdot\rangle)$, where V is a vector space over $\mathbb{R}$ and $\langle\cdot,\cdot\rangle\colon V \times V \to \mathbb{R}$ is a symmetric, bilinear, positive-definite function.

Definition 2.29. Given a finite set S, $\mathbb{R}^S$ will denote the Euclidean space $(\mathbb{R}^S, \langle\cdot,\cdot\rangle)$, where for any f and g members of $\mathbb{R}^S$,

$$\langle f, g\rangle \coloneqq \sum_{i\in S} f(i)g(i).$$

Many notations here should be understood as combinatorial hints. For a given set U, the set $\mathbb{R}^U$ is precisely the functions from U to $\mathbb{R}$, for example. In a "technically correct" spirit, just as much as the 2 in 2^U represents the set $\{0, 1\}$, the number 1 will be used to denote the set $\{0\}$ — or $\{\varnothing\}$, equivalently.

Definition 2.30. Let U and V be finite sets. A *matrix* is a function $A\colon V \times U \to \mathbb{R}$.

In the above definition, the elements of V are called the *rows* of A, and the elements of U are called the *columns* of A.

Definition 2.31. 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

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

Definition 2.32. 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

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

With the set interpretation of 1 in mind, 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.

Do not despair. The just mentioned isomorphism will be properly defined on Section 3.1.

Definition 2.33. 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 2.34. 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. □

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

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

Theorem 2.36. 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,

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

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

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

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

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

Theorem 2.37 (Laplace expansions)**.** Let I and J be finite sets such that $|I| = |J|$. Let $\ell_I \colon I \to [|I|]$ and $\ell_J \colon J \to [|I|]$ be bijective functions. Then

$$\det(A) = \sum_{i \in I} (-1)^{\ell_I(i)+\ell_J(j)} A_{i,j} \det(A[\{i\}^c, \{j\}^c]) \tag{2.38}$$

$$\det(A) = \sum_{j \in J} (-1)^{\ell_I(i)+\ell_J(j)} A_{i,j} \det(A[\{i\}^c, \{j\}^c]). \tag{2.39}$$

Proof. □

Lemma 2.40. 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. □

2.3.2 Projections and direct sum

Definition 2.41. 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 2.42. 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 *projection on* S *along* T is the linear transformation given by

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

Note that if $\mathbb{R}^U = S \oplus T$, then $I = P_{S,T} + P_{T,S}$, which ensures that

$$P_{T,S} = I - P_{S,T} \tag{2.43}$$

Proposition 2.44. Let $P \colon \mathbb{R}^U \to \mathbb{R}^U$. If $P^2 = P$, then P is the projection on $\mathrm{Im}(P)$ along $\mathrm{Null}(P)$.

Proof. □

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 2.45. 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 2.46. Let $S \subseteq \mathbb{R}^U$ be a subspace. Then $P\colon \mathbb{R}^U \to \mathbb{R}^U$ is the orthogonal projection on S if and only if $P^2 = P$ and $P^\mathsf{T} = P$.

Proof. □

2.4 The Laplacian

Definition 2.47. Let $D = (V, A)$ be a digraph. The *head operator of* D is the linear transformation given by

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

Definition 2.48. Let $D = (V, A)$ be a digraph. The *tail operator of* D is the linear transformation given by

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

If a function $f \in \mathbb{R}^A$ is interpreted as a "flow" defined on the arcs, the head operator measures how much flow enters into each vertex per unit of time. Similarly, the tail operator measures how much flow leaves each vertex per unit of time. We can combine both to look at the net result of each vertex.

Definition 2.49. Let $D = (V, A)$ be a digraph. The *divergence operator* is the linear transformation given by

$$B_D \coloneqq H_D - T_D.$$

The matrix B_D will also be called the *incidence matrix*.

Note that, stretching the analogy further, if a function $f \in \mathbb{R}^A$ is interpreted as a "flow" defined on the arcs, the function $(B_D f) \in \mathbb{R}^V$ measures how much flow is accumulated in every vertex, or, equivalently, how "sink-like" every vertex is.

Proposition 2.50. Let $D = (V, A)$ be a digraph. Let $B_D \in \mathbb{R}^{V \times A}$ denote its incidence matrix. If i is any vertex, and a is any arc, then

$$(B_D)_{i,a} = \begin{cases} 1, & \text{if } a = ij \text{ for some } j \in V, \\ -1, & \text{if } a = ji \text{ for some } j \in V, \\ 0, & \text{otherwise.} \end{cases} \tag{2.51}$$

Proof. □

Definition 2.52. Let $D = (V, A)$ be a digraph. The linear transformation given by B_D^T will be called the *gradient operator* of D.

Proposition 2.53. Let $D = (V, A)$ be a digraph. Let $i \in V$ be any vertex. Then

$$B_D^\mathsf{T} = \sum_{ij \in A} e_{ij} \left(e_j^\mathsf{T} - e_i^\mathsf{T} \right).$$

Proof. □

The gradient of a function f on a directed graph is precisely the rate of change in the "directions" given by the directed edges.

Definition 2.54. The *Laplacian* of a weighted digraph $D = (V, A, w)$ is the matrix $L_D \in \mathbb{R}^{V \times V}$ defined as

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

Definition 2.55. Let $G = (V, E, w)$ be any graph, and $D = (V, A, w)$ be an orientation of G. The *Laplacian* of a weighted graph G with respect to D is the matrix $L_G \in \mathbb{R}^{V \times V}$ defined as

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

Proposition 2.56. Let $G = (V, E, w)$ be a weighted graph, and $D = (V, A, w)$ be an orientation of G. Let L_G be the Laplacian of G with respect to D. Then

$$L_G = \sum_{ij \in E} w_{ij}(e_i - e_j)(e_i^\mathsf{T} - e_j^\mathsf{T}) = \sum_{i \in V} e_i \sum_{j \in \delta(i)} w_{ij}(e_i - e_j)^\mathsf{T}.$$

Proof. □

Corollary 2.57. Let $G = (V, E, w)$ be a graph. The Laplacian of G with respect to any orientation D is the same.

Proof. □

Theorem 2.58. Let $G = (V, E, w)$ be a graph, and let $D = (V, A, w_D)$ be such that $ij \in E \iff (ij \in A \text{ and } ji \in A)$ and that $w_D(ij) = w_D(ji) = w(ij)$. Then

$$L_D = L_G.$$

Proof.

$$\begin{aligned}
L_D &= H_D \operatorname{Diag}(w_D) B_D^\mathsf{T} \\
&= H_D \left(\sum_{ij \in A} w_D(ij) e_{ij} (e_j^\mathsf{T} - e_i^\mathsf{T}) \right) \\
&= \left(\sum_{i \in V} e_i \sum_{j \in \delta^{\text{in}}(i)} e_{ji}^\mathsf{T} \right) \left(\sum_{ij \in A} w_D(ij) e_{ij} (e_j^\mathsf{T} - e_i^\mathsf{T}) \right) \\
&= \sum_{i \in V} e_i \sum_{j \in \delta^{\text{in}}(i)} w_D(ji)(e_i^\mathsf{T} - e_j^\mathsf{T}).
\end{aligned}$$

But since $j \in \delta^{\text{in}}(i) \iff j \in \delta(i)$,

$$L_D = \sum_{i \in V} e_i \sum_{j \in \delta(i)} w_D(ij)(e_i^\mathsf{T} - e_j^\mathsf{T}).$$

As $w_D(ij) = w(ij)$, the proof is complete. □

2.5 Harmonic Functions

Definition 2.59. Let $G = (V, E)$ be a graph. A function $f \in \mathbb{R}^V$ is *harmonic* at vertex i if

$$f(i) = \frac{1}{\deg(i)} \sum_{ij \in E} f(j).$$

A vertex where f is not harmonic is called a *pole* of f.

2.6 Measure Theory

Definition 2.60. A *σ-algebra* on a set X is a collection $\Sigma \subseteq 2^X$ such that

(i) the empty set belongs to Σ;

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

(iii) the collection Σ is *closed under countable unions*, that is, if $(E_i)_{i\in\mathbb{N}}$ is a sequence in Σ, then

$$\bigcup_{i\in\mathbb{N}} E_i \in \Sigma.$$

For practical reasons, the set $X \setminus E$ is denoted by E^c.

Theorem 2.61. Let X be a set, and let $\{\Sigma_i : i \in I\}$ be a collection of σ-algebras on X. The collection

$$\Sigma_I := \bigcap_{i\in I} \Sigma_i$$

is a σ-algebra on X.

Proof. By the definition of σ-algebra, the empty set must belong to Σ_i for every $i \in I$. Therefore, it also belongs to Σ_I.

Let $E \in \Sigma_I$. Then, for every $i \in I$, the set E belongs to Σ_i, which implies that $E^c \in \Sigma_i$. But since this holds for every $i \in I$, it follows that $E^c \in \Sigma_I$.

Finally, let $(E_k)_{k\in\mathbb{N}}$ be a sequence of sets in Σ_I. Then, for every $i \in I$ and $k \in \mathbb{N}$, $E_k \in \Sigma_i$. Therefore, for every fixed $i \in I$, the union $\bigcup_{k\in\mathbb{N}} E_k$ is in Σ_i. It follows that

$$\bigcup_{k\in I} E_k \in \Sigma_I.$$ □

Definition 2.62. Let $\mathcal{O} \subseteq 2^X$. Define

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

The collection $\mathcal{O}$ *generates* $\sigma(\mathcal{O})$, and $\sigma(\mathcal{O})$ is called *the σ-algebra generated by* $\mathcal{O}$.

Theorem 2.63. Let $\mathcal{O} \subseteq 2^X$. The collection $\sigma(\mathcal{O})$ is a σ-algebra on X.

Proof. Apply Theorem 2.61. □

Note that the power set of X itself is always a σ-algebra. 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 2.64. Let X be as set. Let $\sigma : 2^{2^X} \to 2^{2^X}$ be the function defined by $\mathcal{O} \mapsto \sigma(\mathcal{O})$.

(i) The function σ is *extensive*, that is, $\mathcal{O} \subseteq \sigma(\mathcal{O})$.

(ii) The function σ is *monotone*, that is, $\mathcal{P} \subseteq \mathcal{O} \implies \sigma(\mathcal{P}) \subseteq \sigma(\mathcal{O})$.

(iii) The function σ is *idempotent*, that is, $\sigma(\sigma(\mathcal{O})) = \sigma(\mathcal{O})$.

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 with a σ-algebra on X such that theorem 2.64 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.

Now, to other practical matters. It must be noted that, just as much as a function must have a domain, a σ-algebra must have a set on which it is built on. It is impossible not to carry around the given set X, just as much as it is impossible not to carry around the domain of a function. Therefore, it is natural to "pack" them together into a single concept. This package is what is called a measurable space.

Definition 2.65. A *measurable space* is an ordered pair (X, Σ) where Σ is a σ-algebra on X.

Definition 2.66. Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. A function $f : X \to Y$ is called *measurable* with respect to (w.r.t.) Σ_X and Σ_Y if the preimage of every measurable set is measurable, that is, if $f^{-1}(\Sigma_Y) \subseteq \Sigma_X$.

Definition 2.67. Let (X, Σ) be a measurable space. A function $\mu : \Sigma \to [0, +\infty]$ is called 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

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

A measurable function f suggests a way to create a measure on (Y, Σ_Y) given a measure on (X, Σ_X). The suggestion is to define

$$\nu := \mu \circ f^{-1}.$$

Therefore, a measurable function is a way to move measures around different measurable spaces. But this procedure demands a measurable space defined on its range. In some cases the function $f : X \to Y$ we want to work with is clear, but the σ-algebra on Y is not. It then becomes desirable to have a way to define a σ-algebra for Y in such a way that f is measurable.

Definition 2.68. A measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space* if $\mathbb{P}(\Omega) = 1$; in this case, $\mathbb{P}$ is called a *probability measure* or *probability distribution*.

Definition 2.69. A *random variable* is a measurable function whose domain is a probability space.

Part I

Laplacian-Based Algorithms

Chapter 3

The Naive Algorithm

The Naive algorithm relies on Kirchhoff's Matrix Tree Theorem, which relies on Cauchy-Binet Formula.

3.1 Cauchy-Binet formula

The main issue we address here is the fact that the determinant is a value defined on linear transformations with same domain and image. We need to extend it, not only to talk about incidence matrices, but even to define the determinant of submatrices. To simply state the Cauchy-Binet Formula formally, it will be necessary to develop new concepts, that actually will play a crucial role in the proof of the Formula.

Lemma 3.1. 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. □

Lemma 3.2. 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 demonstration 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.

Let then $|U| > 0$. Take any $k \in U$, and denote by $U' \coloneqq U \setminus \{k\}$. 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 sum 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:

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

Definition 3.3. 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

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

Proposition 3.4. 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. □

Proposition 3.5. 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

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

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

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

Moreover, if ϕ is bijective,

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

Note that given A and ϕ, 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 a transformation on $\mathbb{R}^V$, and the latter a transformation on $\mathbb{R}^U$. Moreover, it is possible to calculate both determinants, and the next proposition will ensure that both calculations lead to the same result.

Proposition 3.6. 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 3.1 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,

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

Since the mapping $(\sigma \mapsto \phi\sigma\phi^{-1})$ is a bijection from $\mathrm{Sym}(V)$ to $\mathrm{Sym}(U)$,

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

□

The proof above demands a commentary. Let then $f, g\colon U \to V$ be bijective functions. Lemma 3.1 implies that

$$\mathrm{sgn}(f^{-1}g) = \mathrm{sgn}(fg^{-1}).$$

The LHS is the sign of a permutation on V, and the RHS is the sign of a permutation on U. Proposition 3.6 translates this result to a different concept, since for $A \in \mathbb{R}^{V \times U}$ and $\phi\colon V \to U$, we just proved that

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

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

Definition 3.7. 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

$$\det_\phi(A) := \det(AP_\phi).$$

Theorem 3.8. 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

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

Proof. Several previous results come into play. Applying successively, Theorem 2.36, Proposition 3.5, and Proposition 3.6, we have that

$$\det_\phi(A) = \det(AP_\phi) = \det(P_\phi^\top A^\top) = \det(P_{\phi^{-1}} A^\top) = \det(A^\top P_{\phi^{-1}}) = \det_{\phi^{-1}}(A^\top).$$

□

Theorem 3.9. Let U and V be finite sets. Let $f\colon U \to V$ and $g\colon U \to V$ be functions. Then

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

Proof. This proof will be broken into 3 steps:

(1) Prove that if $\det(P_f^\top P_g)$ is nonzero, then both f and g are injective.

(2) Prove that if $\det(P_f^\top P_g)$ is nonzero, then $\mathrm{Im}(f) = \mathrm{Im}(g)$.

(3) Prove that if $\det(P_f^\top P_g)$ is nonzero, then it is equal to $\mathrm{sgn}(f^{-1}g)$.

First, note that if f is not injective, then for any $A \in \mathbb{R}^{V \times U}$, we have that $\det(P_f^\top A) = 0$. 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 $\mathrm{Null}(P_f)$, and, therefore, in $\mathrm{Null}(A^\top P_f)$. This implies that $\det(A^\top P_f) = 0$, and Theorem 2.36 ensures that $\det(P_f^\top A) = 0$.

The contrapositive of this result applied to both $P_f^\top P_g$ and $P_g^\top P_f$ implies step (1).

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

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

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

Into the final step. Let then $f, g\colon U \to V$ 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 $\mathrm{Im}(f) = \mathrm{Im}(g) = S$ and that there exists $g^{-1}\colon S \to U$ inverse of g.

We also have that

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

So that

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

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

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

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 $\mathrm{sgn}(g^{-1}f) = \mathrm{sgn}(f^{-1}g)$. □

The theorem just proved goes further into the direction of relating determinants of linear transformations and signs of permutations. Let $f, g\colon U \to V$ be functions. Note that Proposition 3.5 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 $\mathrm{sgn}(f^{-1}g)$, since both are equal whenever the expression $\mathrm{sgn}(f^{-1}g)$ makes sense, ie, $f^{-1}g$ exists and is invertible.

Proposition 3.10. Let U and V be finite sets. Let $S \subseteq V$. Let $\phi\colon S \to U$ be a bijective function. Then

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

Proof. Proposition 3.5 and Proposition 3.6 ensure that

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

The result then follows from Theorem 3.9, since ϕ is bijective, $\mathrm{Im}(\phi^{-1}) = S$, and

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

Theorem 3.11 (Cauchy-Binet, restricted version)**.** Let U and V be finite sets. Let $f, f\colon U \to V$ 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 3.8 and Proposition 3.10 ensure that

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

Let $S = \operatorname{Im}(f)$. 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,

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

Note that Theorem 3.9 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 theorem above 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 3.12. 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 3.2 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 3.13. 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. Note that suffices to apply Proposition 3.12 twice:

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

Theorem 3.14 (Cauchy-Binet)**.** 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 alternative expression for $\det_{\phi_S}(B[S,U])$.

Theorem 2.36 and Proposition 3.6 allow the manipulations of the matrices, and Proposition 3.5 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 3.13 provides the sum over functions, and once again Theorem 2.36, Proposition 3.6, and Proposition 3.5 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 3.10 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 3.13 to write the product in terms of function matrices, then use Theorem 3.11 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 3.14 is interesting because it reduces the sum of a exponential amount of determinants into a single one. This will be explored first to give a determinant formula to count the amount of spanning trees of a graph, which is exponential. Then, since determinants can be calculated in polynomial time, this will give the first polynomial time algorithm for sampling spanning trees.

3.2 Matrix Tree Theorems

This section is devoted to prove Tutte's Matrix Tree Theorem (3.23), and use it to prove Kirchhoff's result (3.24) as a corollary. Both results arise from an interplay between linear algebra and combinatorics. Therefore, to proper understand the material presented here, we must observe how properties from one world manifest themselves into another. For this reason, most theorems here presented are algebraic conclusions made from combinatorial hypotheses. The results build up into Theorem 3.19 and Proposition 3.22, that actually reach back and give information about the digraph used to define the matrices involved. Combining both results with Theorem 3.14 we obtain Tutte's Matrix Tree Theorem.

Definition 3.15. Let $D = (V, A)$ be a digraph. Let $i, j \in V$. A function $f \in \mathbb{R}^A$ is a *flow* from i to j, or an *ij-flow*, if

$$B_D f \in \mathbb{R}_+(e_j - e_i).$$

Theorem 3.16. Let $D = (V, A)$ be a digraph. Let $G = (V, E)$ be its underlying graph. If there is a walk of positive length in G from vertex i to vertex j, then there exists an ij-flow.

Proof. The proof is by induction on the lenght of the walk. Let $(u_0, \ldots, u_m)$ be a ij-walk.

If $m = 0$, then $x = 0$ suffices.

If $m > 1$, denote by k the vertex u_1 and apply the induction hypothesis to $(u_1, \ldots, u_m)$. Therefore, there is $y \in \mathbb{R}^A$ such that

$$B_D y = e_j - e_k.$$

Either ik or ki is in A. In the former, $x = y + e_{ik}$ is a solution. In the latter, $x = y - e_{ki}$ suffices. □

Theorem 3.16 is a formal statement of the idea that, if there is walk between two vertices, it is possible to establish a flow between them — basically pushing the flow through the walk.

Proposition 3.17. Let $D = (V, A)$ be a digraph. Let $f \in \mathbb{R}^V$. Then $B_D^\mathsf{T} f = 0$ if and only if $f(i) = f(j)$ for every i and j in the same component of the underlying graph of G.

Proof. Let f be constant in every component of G. Let $ij \in A$. Since i and j are on the same component,

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

Since ij was arbitrary, it follows that $B_D^\mathsf{T} f$ is zero.

Now let $f \in \mathbb{R}^V$ be a function such that $B_D^\mathsf{T} f = 0$. Let i and j be distinct vertices on the same component. Then, there is an ij-walk of positive length. Theorem 3.16 ensures that there exists a $x \in \mathbb{R}^A$ such that $B_D x = e_j - e_i$. Therefore

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

Just like in calculus, the gradient of a function being zero is related to it being constant on the connected parts of its domain. From a linear algebra perspective, Proposition 3.17 describres the nullspace of B_D^T according to its underlying graph.

$$\mathrm{Null}(B_D^\mathsf{T}) = \mathrm{span}(\{\, \mathbb{1}_C : C \text{ is a component of } D\text{'s underlying graph.}\}).$$

Proposition 3.18. Let $D = (V, A)$ be a digraph with $|V| - 1$ arcs. Let $i \in V$, and $\phi\colon A \to \{i\}^c$ be a bijective function. If G is not connected, then

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

Proof. Since $\det(B_D[\{i\}^c, A]) = \det(B_D^\mathsf{T}[A, \{i\}^c])$, it suffices to find a nonzero element of $\mathrm{Null}(B_D^\mathsf{T}[A, \{i\}^c])$. Let $C \subsetneq V$ be the component of i. Proposition 3.17 ensures that $B_D^\mathsf{T}\mathbb{1}_{V\setminus C} = 0$.

Note that, since $i \notin V \setminus C$, we have that

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

The proof is done, since $\mathbb{1}_{V\setminus C}$ is nonzero because G is not connected. □

Proposition 3.19. Let $D = (V, A)$ be a digraph with $|V| - 1$ arcs. Let G be its underlying graph. Let $i \in V$. Let $\phi\colon A \to \{i\}^c$ be a bijective function. Then $\det_\phi(B_D[\{i\}^c, A])^2 = [G \text{ is a tree}]$.

Proof. If $\det_\phi(B_D[\{i\}^c, A])^2 = 1$, in particular $\det_\phi(B_D[\{i\}^c, A]) \neq 0$, and the contrapositive of Proposition 3.18 implies that G is connected. Since G has $|V| - 1$ edges, it must be a tree.

Now let G be a tree. If $|V| = 1$, the matrix $B_D[\{i\}^c, A]$ becomes empty. According to our definition, the determinant is an empty product, and therefore the statement holds.

If $|V| > 1$, let j be a vertex distinct from i which is a leaf. Let k be its unique neighbor, and assume that $jk \in A$ — the other case is analogous. Let $\psi\colon A \setminus \{jk\} \to \{i, j\}^c$ be a bijective function. The Laplace expansion along the j-th row, as given in Equation 2.38, ensures that, for some natural r,

$$\det_\phi(B_D[\{i\}^c, A]) = (-1)^r \det_\psi(B_D[\{i, j\}^c, A \setminus \{jk\}]),$$
$$\det_\phi(B_D[\{i\}^c, A])^2 = (-1)^{2r} \det_\psi(B_D[\{i, j\}^c, A \setminus \{jk\}])^2.$$

The induction hypothesis on the graph $G - j$ ensures that the square of the determinant on the right hand side is 1, and concludes the proof. □

Proposition 3.19 is an algebraic criterion to determine if the underlying graph of a digraph is a tree. It bring us close to characterize arborescences. We continue on this path with the following theorem.

Proposition 3.20. Let $D = (V, A)$ be a digraph with $|V| - 1$ arcs. Let $i \in V$. Let $\phi\colon A \to \{i\}^c$ be a bijective function. Then $\det_\phi(H_D[\{i\}^c, A]) \neq 0$ implies that

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

Proof. If j is a vertex with indegree 0, then e_j is in $\mathrm{Null}(H_D^\mathsf{T})$, and $\det(H_D[\{i\}^c, A]) = 0$.

Therefore, if the determinant is nonzero, every vertex different from i has indegree at least 1. But since there are $|V| - 1$ arcs, every vertex different from i has indegree precisely 1. □

Proposition 3.21. Let $D = (V, A)$ be an i-arborescence. Let $\phi\colon A \to \{i\}^c$ be a bijective function. Then

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

Proof. Let $D = (V, A)$ be a minimal counterexample, ie, a minimal i-arborescence such that both determinants differ. It is impossible for $|V|$ to be one, since in such case $B_D = H_D = [0]$.

Theorem 2.27 ensures there is a vertex distinct from i with outdegree zero. Let j be it. The row corresponding to j in $H_D[\{i\}^c, A]$ has only one nonzero entry, and its value is precisely 1. Let $kj \in A$ be the corresponding entry. Also, let $\psi\colon A \setminus \{kj\} \to \{i, j\}^c$ be a bijective function. Then the Laplace expansion along the j-th row, as given in Equation 2.38, ensures that

$$\det_\phi(H_D[\{i\}^c, A]) = \det_\psi(H_D[\{i, j\}^c, A \setminus \{kj\}]).$$

The same argument ensures that

$$\det_\phi(B_D[\{i\}^c, A]) = \det_\psi(B_D[\{i, j\}^c, A \setminus \{kj\}]).$$

But this implies that the graph $D - j$ is also a counterexample, contradicting the minimality of D. □

Proposition 3.22. Let $D = (V, A)$ be a digraph with $|V| - 1$ arcs. Let $i \in V$. Let $\phi\colon A \to \{i\}^c$ be a bijective function. Then

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

Proof. If (V, S) is not an i-arborescence, then Theorem 3.19 and Proposition 3.20 ensure that the product is zero.

Therefore the determinant is nonzero only when (V, S) is an i-arborescence. But in such case, Theorem 3.19 and Proposition 3.21 ensures that the product is one. □

Theorem 3.23 (Tutte's Matrix Tree Theorem)**.** Let $D = (V, A, w)$ be a weighted digraph. Let $i \in V$. Then

$$\det(L_D[\{i\}^c, \{i\}^c]) = \sum_{F \in \mathcal{T}_D(i)} \prod_{e \in F} w_e.$$

Proof. For every $S \in \binom{A}{|V|-1}$, let $\phi_S\colon S \to \{i\}^c$ be bijective functions. First note that Theorem 3.14 ensures that, for every fixed $S \in \binom{A}{|V|-1}$, we have

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

Moreover, Theorem 2.34 and Theorem 3.14 ensure that

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

Proposition 3.22 is used in the change of summation index. □

Theorem 3.24 (Kirchhoff's Matrix Tree Theorem)**.** Let L be the laplacian of a weighted graph $G = (V, E, \psi, w)$. Let $i \in V$. Then

$$\det(L[\{i\}^c, \{i\}^c]) = \sum_{F \in \mathcal{T}_G} \prod_{e \in F} w_e.$$

Proof. Let $G = (V, E, w)$ be weighted graph. Let $D = (V, A, w_D)$ be such that

$$ij \in E \iff (ij \in A \text{ and } ji \in A),$$

and $w_D(ij) = w_D(ji) = w(ij)$. Then Theorem 2.58 ensures that $L_G = L_D$. Therefore, Theorem 3.23, implies that

$$\det(L_G[\{i\}^c, \{i\}^c]) = \det(L_D[\{i\}^c, \{i\}^c]) = \sum_{S \in \mathcal{T}_D(i)} \prod_{e \in S} w_D(e).$$

Note that, by construction of D, for every $S \in \mathcal{T}_D(i)$, if $F \subseteq E$ is such that (V, F) is the underlying graph of (V, S), then $\prod_{e \in F} w(e) = \prod_{e \in S} w_D(e)$. Therefore, remains only to change the summation index.

In other words, to prove the statement suffices to prove that for every $T \in \mathcal{T}_G$ and for every $i \in V$, there is a unique $F \in \mathcal{T}_D(i)$ such that the underlying graph of (V, F) is (V, T).

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

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

Let then $|V| > 1$. Let $T \in \mathcal{T}_G$ and let j be a leaf of (V, T) which is distinct from i. Let k be the only vertex adjacent to j, ie, let $kj \in T$. Note that $T \setminus \{kj\} \in \mathcal{T}_{G[\{j\}^c]}$. Therefore, the induction hypothesis applies, and there is a unique $F' \in \mathcal{T}_{D[\{j\}^c]}$ such that $(\{j\}^c, T \setminus \{kj\})$ is the underlying graph of $(\{j\}^c, F')$. Then, $F := \{kj\} \cup F'$ is an i-arborescence of D whose underlying graph is (V, T). Note that, since $j \neq i$, then we have that $kj \in F$ in every i-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. □

3.3 The Algorithm

Definition 3.25. Let $D = (V, A, \psi, w)$ be a weighted digraph, and let $i \in V$. Denote

$$\Phi(D, i) := \det(L_D[\{i\}^c, \{i\}^c]).$$

Proposition 3.26. Let $D = (V, A, \psi, w)$ be a weighted digraph, and let $i \in V$. Then

$$\Phi(D, i) = \sum_{T \in \mathcal{T}_D(i)} \prod_{a \in T} w(a).$$

Proof. Apply Theorem 3.23 to the definition of $\Phi(D, i)$. □

The following propositions serve to 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 algorithm, and at its inductive proof of correctness.

Proposition 3.27. Let $D = (V, A, \psi, w)$ be a weighted digraph, let $i \in V$ and let $a_0 \in \delta^{\text{out}}(i)$ not be a loop. Let $S := \{\, T \in \mathcal{T}_D(i) : a_0 \in T \}$. Then $\phi \colon \mathcal{T}_{D/a_0}(a_0) \to S$ given by $\phi := (T \;\mapsto\; T \cup \{a_0\})$ is bijective.

Proof. Let $f \colon V \to V \setminus \{i, j\} \cup \{a_0\}$ be defined as

$$\begin{aligned} f|_{V \setminus \{i,j\}} &= (x \mapsto x), \\ f(i) &= a_0, \\ f(j) &= a_0. \end{aligned}$$

Moreover, define $\hat{f} := (ij \;\mapsto\; f(i)f(j))$. Then, according to the definition of contraction, we have that

$$D/a_0 = (f(V), A \setminus \{a_0\}, \hat{f}\psi, w|_{A \setminus \{a_0\}}).$$

Let then $a \in A \setminus \{a_0\}$ be any arc. Note that $\hat{f}\psi(a) = a_0 k$ if and only if $\psi(a) \in \{ik, jk\}$. This, along with the fact that $\left|\delta^{\text{in}}(a_0)\right| = 0$ for every $T \in \mathcal{T}_{D/a_0}(a_0)$, ensures that $\left|\delta^{\text{in}}(i)\right| = \left|\delta^{\text{in}}(j)\right| = 0$ in (V, T). We can then conclude that the graph $(V, T \cup \{a_0\})$ satisfies the indegree condition for being an i-arborescence, i.e., that $\left|\delta^{\text{in}}(j)\right| = [i \neq j]$.

Note that since $\psi(a_0) = ij$ for distinct i and j, we have both that $|T| = |V| - 2$ and $|f(V)| = |V| - 1$. Therefore, to finish the proof, suffices to show that given $T \subseteq A \setminus \{a_0\}$, the digraph $(f(V), T, \hat{f}\psi)$ has a connected underlying subgraph if and only if $(V, T \cup \{a_0\}, \psi)$ has a connected underlying subgraph. We proceed in this direction.

Suppose $T \subseteq A \setminus \{a_0\}$ is such that $(f(V), T, \hat{f}\psi)$ has a connected underlying subgraph. Therefore, for every vertex k in $f(V) \setminus \{a_0\}$, there is a walk, in the underlying graph, between a_0 and k. Let a be the first arc in this walk.

Note that a is such that either $\hat{f}\psi(a) \in \{a_0 l, l a_0\}$, for some vertex l. Either way, the underlying graph has an edge il or an edge jl. Therefore, there is either a walk from i to k or from j to k in the underlying graph of (V, T, ψ). Since $\psi(a_0) = ij$, it follows that $(V, T \cup \{a_0\}, \psi)$ has a connected underlying subgraph.

Suppose now that $(V, T \cup \{a_0\}, \psi)$ has a connected underlying subgraph. Note then that the underlying graph of (V, T, ψ) must have two components, such that one contains the vertex i and the other contains the vertex j. Therefore, in the underlying graph, for every vertex $k \in V \setminus \{i, j\}$, there must be either a walk from i to k, or a walk from j to k. Th proof is finished, since this implies that there is a walk from a_0 to k in the underlying graph of $(f(V), T, \hat{f}\psi)$. □

Proposition 3.28. Let $D = (V, A, \psi, w)$ be a weighted digraph, let i be a vertex in it, and let $a_0 \in \delta^{\text{out}}(i)$ not be a loop. Then

$$\Phi(D, i) = w(a)\, \Phi(D/a_0, a_0) + \Phi(D - a_0, i).$$

Proof. Let $T \in \mathcal{T}_D(i)$. Note that,

$$1 = [a_0 \in T] + [a_0 \notin T].`$$

This observation, along with Theorem 3.24 and Propositions 3.27, ensure that

$$\begin{aligned}
\det(L_G[\{i\}^c, \{i\}^c]) &= \sum_{T \in \mathcal{T}_D(i)} \prod_{e \in T} w(e) \\
&= \sum_{T \in \mathcal{T}_D(i)} ([a \in T] + [a \notin T]) \prod_{e \in T} w(e) \\
&= \sum_{T \in \mathcal{T}_D(i)} [a \in T] \prod_{e \in T} w(e) + \sum_{T \in \mathcal{T}_D(i)} [a \notin T] \prod_{e \in T} w(e) \\
&= w(a) \left(\sum_{T \in \mathcal{T}_D(i)} [a \in T] \prod_{e \in T \setminus \{a\}} w(e) \right) + \left(\sum_{T \in \mathcal{T}_{D-a}} \prod_{e \in T} w(e) \right) \\
&= w(a) \left(\sum_{T \in \mathcal{T}_{D/a}} \prod_{e \in T} w(e) \right) + \left(\sum_{T \in \mathcal{T}_{D-a}} \prod_{e \in T} w(e) \right) \\
&= w(a) \det(L_{D/a}[\{i\}^c, \{i\}^c]) + \det(L_{D-a}[\{i\}^c, \{i\}^c]).
\end{aligned}$$

□

Loops are indeed a special case. Given a weighted digraph $D = (V, A, \psi, w)$, for any vertex $i \in V$ and loop $a_0 \in \delta^{\text{out}}(i)$, note that the function that maps i 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, so that

$$\Phi(D, i) = \Phi(D - a_0, i) = \Phi(D/a_0, a_0),$$

contrary to Proposition 3.28.

Loops aside, Proposition 3.28 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 restriction, not to fight against it. The algorithm will actually be defined as a random variable, whose definition relies on a suitable "randomness source". This is akin to how a programmer could just call a `rand` function. In that context, there is little interest on what `rand` does. In this context, there is no *immediate* interest on the random variable that poses as "randomness source", since it is a purely measure theoretical structure.

Definition 3.29. Let A be a finite set. A *finite randomness source* is a set of measurable functions $X_a : \Omega \to [0, 1]$ such that for every $a \in A$ and for every interval (b, c) in $[0, 1]$, we have that

$$\mathbb{P}(X_a \in (b, c)) = c - b,$$

and such that, for distinct $a_0, a_1 \in A$ and measurable sets $R, S \subseteq [0, 1]$, we have that

$$\mathbb{P}(a_0 \in R, a_1 \in S) = \mathbb{P}(a_0 \in R) \cdot \mathbb{P}(a_1 \in S).$$

However, 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. This aditional information is a total order in it. Note that to require a total order on the set of arcs is by no means a restriction.

Finally, before going to the definition, a last 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 outputs the first which satisfies the described condition.

Definition 3.30. Let $D = (V, A, \psi, w)$ be a weighted graph. Let $i \in V$ be a vertex in it. Let $(A, \leq)$ be a totally ordered set. The *naive algorithm* is the function $\mathcal{A}(D, i)\colon \Omega \to \mathcal{T}_D(i) \cup \{\perp\}$ defined as

$$\mathcal{A}(D,i)(\omega) \coloneqq \begin{cases} \perp, & \text{if } \Phi(D,i) = 0, \quad (\texttt{error case}) \\ \varnothing, & \text{if } \delta^{\text{out}}(i) = \varnothing, \quad (\texttt{base case}) \\ \mathcal{A}(D-a, i), & \text{if } X_a(\omega) \leq \frac{\Phi(D-a,i)}{\Phi(D,i)}, \quad (\texttt{drop case}) \\ \mathcal{A}(D/a, a) \cup \{a\}, & \text{otherwise.} \quad (\texttt{take case}) \end{cases}$$

where $a \coloneqq \min \delta^{\text{out}}(i)$.

The appearence of $\perp$ on the definition reflects the fact that it is possible for the algorithm to receive a digraph with many arborescences and to fail to output one of them. This will not actually be a problem, but for now, $\perp$ must be carried around.

Theorem 3.31. Let D be a weighted digraph, and let i be a vertex in it. Then

1. The function $\mathcal{A}(D, i)$ is a random variable;
2. For every $T \in \mathcal{T}_D(i)$, it holds that

$$\mathbb{P}(\mathcal{A}(D,i) = T) = \frac{1}{\Phi(D,i)} \prod_{e \in T} w(e);$$

3. If D has at least one i-arborescence, then $\mathbb{P}(\mathcal{A}(D, i) = \perp) = 0$.

Proof. First, note that if D has no i-arborescences, then $\Phi(D, i) = 0$. Therefore,

$$\Omega = \{\mathcal{A}(D,i) = \perp\},$$

since the algorithm will always go through `error case`, regardless of the input ω. As the measurable sets form a σ-algebra, it follows that $\{\mathcal{A}(D, i) = \perp\}$ is measurable. In such a case, (1), (2) and (3) hold.

Let then $D = (V, A, \psi, w)$ be a weighted digraph with at least one i-arborescence. We proceed by proving that for every $T \in \mathcal{T}_D(i)$,

(i) The set $\{\mathcal{A}(D, i) = T\}$ is measurable, and

(ii) $\mathbb{P}(\mathcal{A}(D, i) = T) = \frac{1}{\Phi(D,i)} \prod_{e \in T} w(e)$.

We proceed by induction on $|A|$.

If $A = \varnothing$, since $\mathcal{T}_D(i)$ is not empty, it follows that $\mathcal{T}_D(i) = \{\varnothing\}$. In such a case, note that

$$\Omega = \{\mathcal{A}(D,i) = \varnothing\},$$

since the naive algorithm will output $\varnothing$ regardless of the input $\omega \in \Omega$. Since the measurable sets are a σ-algebra, this implies that $\{\mathcal{A}(D, i) = \varnothing\}$ is measurable. Also, since it is a probability space, its measure must be 1, which is equal to the RHS of (ii), since $\Phi(D, i) = 1$ and the empty product "collapses" into 1.

Let then A be nonempty, and let $T \in \mathcal{T}_D(i)$ be any i-arborescence. Since $T \in \mathcal{T}_D(i)$ and $w > 0$, it follows that $\Phi(D, i) > 0$ and that $\delta^{\text{out}}(i)$ is not empty. In such a case, let a be the minimum element of $\delta^{\text{out}}(i)$.

Observe that we can now assume that the algorithm is not on `error case`, since $\Phi(D, i) > 0$, nor on `base case`, since $\delta^{\text{out}}(i) \neq \varnothing$.

Note that if a is a loop, then

$$\Phi(D,i) = \Phi(D-a,i),$$

so that the simple fact that $X_a(\omega) \leq 1$ for every $\omega \in \Omega$ implies that the algorithm is in `drop case`. The inductive hypothesis then ensure that (i) and (ii) hold.

If a is not a loop, there are two cases to consider, depending on whether a is in T or not.

If $a \in T$, then the set $\{\mathcal{A}(D,i) = T\}$ is a subset of $\{a \in \mathcal{A}(D,i)\}$. But for every $\omega \in \Omega$ such that $a \in \mathcal{A}(D,i)(\omega)$, we are dealing with `take case`, so that

$$\{\mathcal{A}(D,i) = T\} = \{\mathcal{A}(D,i) = T, a \in \mathcal{A}(D,i)\} = \left\{\mathcal{A}(D/a,a) = T \setminus \{a\}, \frac{\Phi(D-a,i)}{\Phi(D,i)} < X_a\right\}.$$

The last set in the above equation is measurable because it is the intersection of two measurable sets. The first, has its measurability ensured by the induction hypothesis, and the second, by the fact that $X_a \colon \Omega \to [0,1]$ is measurable. Note that Proposition 3.27 is used to ensure that $T \setminus \{a\}$ is a a-arborescence in D/a.

We then have that

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(D,i) = T) &= \mathbb{P}\left(\mathcal{A}(D/a,a) = T \setminus \{a\}, \frac{\Phi(D-a,i)}{\Phi(D,i)} < X_a\right) \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\}) \cdot \mathbb{P}\left(\frac{\Phi(D-a,i)}{\Phi(D,i)} < X_a\right) && \text{since the } X_a \text{ are independent} \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\}) \cdot \left(1 - \frac{\Phi(D-a,i)}{\Phi(D,i)}\right) \\
&= \mathbb{P}(\mathcal{A}(D/a,a) = T \setminus \{a\}) \cdot w(a)\frac{\Phi(D/a,a)}{\Phi(D,i)} && \text{by Proposition 3.28} \\
&= \left(\frac{1}{\Phi(D/a,a)} \prod_{e \in T \setminus \{a\}} w(e)\right) \cdot w(a)\frac{\Phi(D/a,a)}{\Phi(D,i)} && \text{by induction hypothesis} \\
&= \frac{1}{\Phi(D,i)} \prod_{e \in T} w(e).
\end{aligned}$$

If $a \notin T$, then the set $\{\mathcal{A}(D,i) = T\}$ is a subset of $\{a \notin \mathcal{A}(D,i)\}$. But for every $\omega \in \Omega$ such that $a \notin \mathcal{A}(D,i)(\omega)$, we are dealing with `drop case`, so that

$$\{\mathcal{A}(D,i) = T\} = \{\mathcal{A}(D,i) = T, a \notin \mathcal{A}(D,i)\} = \left\{\mathcal{A}(D-a,i) = T, X_a \leq \frac{\Phi(D-a,i)}{\Phi(D,i)}\right\}.$$

The last set on the above equation is the intersection of two measurable sets, the first with its measurability assured by the induction hypothesis, and the second because $X_a \colon \Omega \to [0,1]$ is a measurable.

Therefore,

$$\begin{aligned}
\mathbb{P}(\mathcal{A}(D,i) = T) &= \mathbb{P}\left(\mathcal{A}(D-a,i) = T, X_a \leq \frac{\Phi(D-a,i)}{\Phi(D,i)}\right) \\
&= \mathbb{P}(\mathcal{A}(D-a,i) = T) \cdot \mathbb{P}\left(X_a \leq \frac{\Phi(D-a,i)}{\Phi(D,i)}\right) && \text{since the } X_a \text{ are independent} \\
&= \mathbb{P}(\mathcal{A}(D-a,i) = T) \cdot \frac{\Phi(D-a,i)}{\Phi(D,i)} \\
&= \left(\frac{1}{\Phi(D-a,i)} \prod_{e \in T} w(e)\right) \cdot \frac{\Phi(D-a,i)}{\Phi(D,i)} && \text{by induction hypothesis} \\
&= \frac{1}{\Phi(D,i)} \prod_{e \in T} w(e).
\end{aligned}$$

This finishes the proof of (i) and (ii). We know focus back in the original statement of the theorem. We have only showed that the pre image of every arborescence is measurable. Remains to show that $\mathcal{A}(D,i)$ is indeed a random variable (1), and that (3) holds.

To show that $\mathcal{A}(D,i)$ is a random variable, remains only to show that the preimage of $\perp$ is measurable. Note then that

$$\{\mathcal{A}(D,i) = \perp\} = \Omega \setminus \left(\bigcup_{T \in \mathcal{T}_D(i)} \{\mathcal{A}(D,i) = T\}\right),$$

and the RHS is the complement of a finite union of measurable sets, so that it is indeed measurable.

Finally, using Proposition 3.26, we have that

$$\begin{aligned}\mathbb{P}(\mathcal{A}(D,i)=\perp) &= 1-\sum_{T\in\mathcal{T}_D(i)}\mathbb{P}(\mathcal{A}(D,i)=T)\\ &= 1-\left(\frac{1}{\Phi(D,i)}\sum_{T\in\mathcal{T}_D(i)}\prod_{e\in T}w(e)\right)\\ &= 1-\left(\frac{\Phi(D,i)}{\Phi(D,i)}\right)=0,\end{aligned}$$

which demonstrates (3) and finishes the proof. □

If we let $\perp$ represent a division by zero error, the pseudocode for the algorithm can be written as follows:

function Sample(D, i)
  Let $D = (V, A, \psi, w)$.
  if $\delta^{\text{out}}(i) = \varnothing$ **then return** $\varnothing$.
  Let $a \in \delta^{\text{out}}(i)$, and let $p \leftarrow \Phi(D-a, i)/\,\Phi(D, i)$.
  Let x be a uniform random variable in the interval $[0, 1]$.
  if $x \leq p$ **then return** Sample$(D - a, i)$.
  else **return** $\{a\} \cup$ Sample$(D/a, a)$.

Chapter 4

The Harvey-Xu Algorithm

4.1 The Moore-Penrose pseudoinverse

Remember that, for a given linear subspace $S \subseteq \mathbb{R}^U$, the matrix P_S is the unique orthogonal projection along S.

This section is based on the results discussed on projections and on the following result.

Lemma 4.1. Let U be a finite set. Let $A\colon \mathbb{R}^U \to \mathbb{R}^U$. Then

(a) $A = P_{\mathrm{Im}(A)}A$,

(b) $A = AP_{\mathrm{Null}(A)^\perp}$.

Proof. To prove (a), suffices to note that for every $x \in \mathbb{R}^U$, we have that $Ax \in \mathrm{Im}(A)$, so that $P_{\mathrm{Im}(A)}Ax = Ax$.

To prove (b), note that $\mathbb{R}^U = \mathrm{Null}(A) \oplus \mathrm{Null}(A)^\perp$. Therefore, for every $x \in \mathbb{R}^U$, there are unique $(y, z) \in \mathrm{Null}(A) \times \mathrm{Null}(A)^\perp$ such that $x = y + z$. Then

$$Ax = A(y + z) = Ay + Az = Az = AP_{\mathrm{Null}(A)^\perp}x. \qquad \square$$

Definition 4.2 (Moore-Penrose pseudoinverse)**.** Let $A\colon \mathbb{R}^U \to \mathbb{R}^V$. A *(Moore-Penrose) pseudoinverse* is a linear transformation $A^\dagger\colon \mathbb{R}^V \to \mathbb{R}^U$ such that

(a) $AA^\dagger = P_{\mathrm{Im}(A)}$,

(b) $A^\dagger A = P_{\mathrm{Im}(A^\dagger)}$.

Note that the definition of the pseudoinverse is symmetric on A and $A^\dagger$, so that A is the pseudoinverse of $A^\dagger$. Also, properties (a) and (b) of the definition, together with Lemma 4.1 imply

$$AA^\dagger A = P_{\mathrm{Im}(A)}A = A, \tag{4.3}$$

$$A^\dagger AA^\dagger = P_{\mathrm{Im}(A^\dagger)}A^\dagger = A^\dagger. \tag{4.4}$$

Proposition 4.5. Let $A\colon \mathbb{R}^U \to \mathbb{R}^V$, and let $A^\dagger\colon \mathbb{R}^V \to \mathbb{R}^U$ be a pseudoinverse of A. Then

(a) $\mathrm{Null}(A^\dagger) = \mathrm{Im}(A)^\perp$,

(b) $\mathrm{Im}(A^\dagger) = \mathrm{Null}(A)^\perp$.

Proof. It is possible to conclude that $\mathrm{Null}(AA^\dagger) = \mathrm{Null}(A^\dagger)$ from this seemingly pointless computation

$$\mathrm{Null}(A^\dagger) \subseteq \mathrm{Null}(AA^\dagger) \subseteq \mathrm{Null}(A^\dagger AA^\dagger) = \mathrm{Null}(A^\dagger).$$

But this implies (a), since $AA^\dagger = P_{\mathrm{Im}(A)}$, so that the nullspace of $AA^\dagger$ is $\mathrm{Im}(A)^\perp$.

Similarly,

$$\mathrm{Null}(A) \subseteq \mathrm{Null}(A^\dagger A) \subseteq \mathrm{Null}(AA^\dagger A) = \mathrm{Null}(A)$$

ensures that $\mathrm{Null}(A) = \mathrm{Null}(A^\dagger A)$, which implies (b), since $\mathrm{Null}(A^\dagger A) = \mathrm{Null}(P_{\mathrm{Im}(A^\dagger)}) = \mathrm{Im}(A^\dagger)^\perp$. $\square$

Proposition 4.6. Let $A\colon \mathbb{R}^U \to \mathbb{R}^V$. If A has a pseudoinverse, it is unique.

Proof. Let B and C be two pseudoinverses of A. Proposition 4.5 implies that $\mathrm{Im}(B) = \mathrm{Im}(C)$, so that

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

We recall that a linear transformation is injective if and only if its nullspace is $\{0\}$. Therefore, for any linear transformation $A\colon \mathbb{R}^U \to \mathbb{R}^V$, its restriction on $\mathrm{Null}(A)^\perp$ is injective. It is then possible to define its inverse,

$$\left(A\big|_{\mathrm{Null}(A)^\perp}\right)^{-1} \colon \mathrm{Im}(A) \to \mathrm{Null}(A)^\perp.$$

Since it is the inverse of a linear transformation, it is a linear transformation, and we have that

$$\left(A\big|_{\mathrm{Null}(A)^\perp}\right)\left(A\big|_{\mathrm{Null}(A)^\perp}\right)^{-1} = I_{\mathrm{Im}(A)}. \tag{4.7}$$

$$\left(A\big|_{\mathrm{Null}(A)^\perp}\right)^{-1}\left(A\big|_{\mathrm{Null}(A)^\perp}\right) = I_{\mathrm{Null}(A)^\perp}. \tag{4.8}$$

This procedure is general, and allows us to prove the existence of the pseudoinverse.

Proposition 4.9. Let $A\colon \mathbb{R}^U \to \mathbb{R}^V$ be any linear transformation. A has a pseudoinverse, and it holds that

$$A^\dagger = \left(A\big|_{\mathrm{Null}(A)^\perp}\right)^{-1} P_{\mathrm{Im}(A)}.$$

Proof. Since the RHS of the equation is defined for every A, and since the pseudoinverse is unique, suffices to prove that the RHS is a pseudoinverse of A. Denote by B the matrix $\left(A\big|_{\mathrm{Null}(A)^\perp}\right)^{-1}$. Since B is surjective in $\mathrm{Null}(A)^\perp$ and $P_{\mathrm{Im}(A)}$ is surjective in $\mathrm{Im}(A)$, we have that

$$\mathrm{Im}(BP_{\mathrm{Im}(A)}) = \mathrm{Im}(B) = \mathrm{Null}(A)^\perp.$$

Lemma 4.1 and Equation 4.7 are the only tools necessary here:

$$\begin{aligned} BP_{\mathrm{Im}(A)}A = BAP_{\mathrm{Null}(A)^\perp} = B(A|_{\mathrm{Null}(A)^\perp})P_{\mathrm{Null}(A)^\perp} = I_{\mathrm{Null}(A)^\perp}P_{\mathrm{Null}(A)^\perp} = P_{\mathrm{Null}(A)^\perp}, \\ ABP_{\mathrm{Im}(A)} = (A|_{\mathrm{Null}(A)^\perp})BP_{\mathrm{Im}(A)} = I_{\mathrm{Im}(A)}P_{\mathrm{Im}(A)} = P_{\mathrm{Im}(A)}. \end{aligned} \qquad \square$$

Theorem 4.10. Let $A \in \mathbb{R}^{V\times U}$. Then it holds

(a) $\mathrm{Im}(A^\top A) = \mathrm{Im}(A^\top)$, and

(b) $\mathrm{Null}(A^\top A) = \mathrm{Null}(A)$.

Proof. To say that $\mathrm{Im}(A) = \mathrm{Null}(A^\top)^\perp$ is to say that A is surjective in $\mathrm{Null}(A^\top)^\perp$, so that $\mathrm{Im}(A^\top A) = \mathrm{Im}(A)$.

Similarly, to say that $\mathrm{Null}(A^\top) = \mathrm{Im}(A)^\perp$ is to say that $A^\top$ is injetive when restricted to $\mathrm{Im}(A)$, so that $\mathrm{Null}(A^\top A) = \mathrm{Null}(A)$. $\square$

Proposition 4.11. Let $A\colon \mathbb{R}^U \to \mathbb{R}^V$ be any linear transformation. Then

$$(A^\top)^\dagger = (A^\dagger)^\top.$$

Proof. First, note that $\mathrm{Im}(A^\dagger) = \mathrm{Null}(A)^\perp = \mathrm{Im}(A^\top)$. Moreover, $\mathrm{Im}(A) = \mathrm{Null}(A^\dagger) = \mathrm{Im}((A^\dagger)^\top)$.

We can then show that $(A^\dagger)^\top$ satisfies both conditions on the pseudoinverse definition of $A^\top$:

$$\begin{aligned} A^\top(A^\dagger)^\top = (A^\dagger A)^\top = A^\dagger A = P_{\mathrm{Im}(A^\dagger)} = P_{\mathrm{Im}(A^\top)}, \\ (A^\dagger)^\top A^\top = (AA^\dagger)^\top = AA^\dagger = P_{\mathrm{Im}(A)} = P_{\mathrm{Im}((A^\dagger)^\top)}. \end{aligned} \qquad \square$$

Proposition 4.12. Let $A \in \mathbb{R}^{V\times U}$. Then it holds that

(a) $A^\dagger = (A^\top A)^\dagger A^\top$.

(b) $A^\dagger = A^\intercal (AA^\intercal)^\dagger$.

Proof. Let $B = (A^\intercal A)^\dagger A^\intercal$. First note that AB is a projection:

$$ABAB = A(A^\intercal A)^\dagger A^\intercal A(A^\intercal A)^\dagger A^\intercal = A(A^\intercal A)^\dagger A^\intercal = AB.$$

Also, Proposition 4.11 ensures that AB is orthogonal, since

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

Remains to show that $\mathrm{Im}(AB) = \mathrm{Im}(A)$. Note that Lemma 4.1 and Theorem 4.10 imply that

$$A = AP_{\mathrm{Null}(A)^\perp} = AP_{\mathrm{Im}(A^\intercal A)} = AP_{\mathrm{Im}((A^\intercal A)^\dagger)} = A(A^\intercal A)^\dagger A^\intercal A,$$

This is enough for the image equality, since

$$\mathrm{Im}(A) = \mathrm{Im}(A(A^\intercal A)^\dagger A^\intercal A) \subseteq \mathrm{Im}(A(A^\intercal A)^\dagger A^\intercal) = \mathrm{Im}(AB) \subseteq \mathrm{Im}(A).$$

Therefore, $\mathrm{Im}(A) = \mathrm{Im}(AB)$, so that AB is the orthogonal projection on $\mathrm{Im}(A)$. Remains to show that BA is the orthogonal projection on $\mathrm{Im}(A^\dagger)$. But this is quite simpler: just note that the definition of the pseudoinverse of $A^\intercal A$, Theorem 4.10 and Proposition 4.5 imply that

$$BA = (A^\intercal A)^\dagger A^\intercal A = P_{\mathrm{Im}(A^\intercal A)} = P_{\mathrm{Im}(A^\intercal)} = P_{\mathrm{Im}(A^\dagger)}.$$

This finishes the proof of (a). To prove (b), just note that (a) applied to the transpose of A ensures that

$$(A^\dagger)^\intercal = (A^\intercal)^\dagger = (AA^\intercal)^\dagger A,$$

and then, applying Proposition 4.11, we have that

$$A^\dagger = ((AA^\intercal)^\dagger A)^\intercal = A^\intercal (AA^\intercal)^\dagger.$$

□

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 4.13. Let $A \in \mathbb{R}^{V \times U}$ and $B \in \mathbb{R}^{U \times T}$. If A has full column rank and B has full row rank, then

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

Proof. To say that A has full column rank is the same as to say that A is injective, ie, that $\mathrm{Null}(A) = \{0\}$. But then, Theorem 4.10 and Theorem **??** gives that

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

Therefore, $A^\intercal A$ is surjective. The same reasoning gives that

$$\mathrm{Null}(A^\intercal A) = \mathrm{Null}(A) = \{0\}.$$

Therefore, $A^\intercal A$ is injective, and, then, invertible.

Since B has full row rank, $B^\intercal$ has full column rank, and the argument above ensures that $BB^\intercal$ is also invertible. Applying the convenient equations from Proposition 4.12, it follows that

$$B^\dagger A^\dagger = B^\intercal (BB^\intercal)^{-1} (A^\intercal A)^{-1} A^\intercal.$$

First, note that $ABB^\dagger A^\dagger = A(A^\intercal A)^{-1} A^\intercal$, a matrix product that is both orthogonal, and also a projection. Moreover, using the fact that $(A^\intercal A)^{-1}$ is invertible, Theorem 4.10 and the fact that B is surjective, we have that

$$\mathrm{Im}(A(A^\intercal A)^{-1} A^\intercal) = \mathrm{Im}(AA^\intercal) = \mathrm{Im}(A) = \mathrm{Im}(AB).$$

Similarly, observe that $B^\dagger A^\dagger AB = B^\intercal (BB^\intercal)^{-1} B$, a matrix product that is both orthogonal and also a projection. Moreover, using the fact that $(BB^\intercal)^{-1}$ is invertible, Theorem 4.10 and the fact that A is injective, Theorem **??** and Proposition 4.5, we have that

$$\mathrm{Im}(B^\intercal (BB^\intercal)^{-1} B) = \mathrm{Im}(B^\intercal B) = \mathrm{Im}(B^\intercal) = \mathrm{Null}(B)^\perp = \mathrm{Null}(AB)^\perp = \mathrm{Im}((AB)^\dagger).$$

□

We finish this section by actually calculating the pseudoinverse of a matrix that will be useful in our next section.

Example 4.14. Let V be a finite set, let $i \in V$, and let $A \in \mathbb{R}^{V \times \{i\}^c}$ be given by

$$A[\{i\}^c, \{i\}^c] = I,$$
$$A[i, \{i\}^c] = -\mathbb{1}^\mathsf{T}.$$

In other words, assuming the first rows are indexed by $\{i\}^c$, let

$$A = \begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}. \tag{4.15}$$

We wish to calculate $A^\dagger$. First, note that for $x \in \mathbb{R}^{\{i\}^c}$,

$$Ax = 0 \iff \begin{bmatrix} x \\ -\mathbb{1}^\mathsf{T} x \end{bmatrix} = 0 \iff x = 0.$$

So that $\operatorname{Null}(A) = \{0\}$.

Moreover, for $y \in \mathbb{R}^V$, we have that $y \perp \mathbb{1}$ if and only if there is $x \in \mathbb{R}^{\{i\}^c}$ such that

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

So that $\operatorname{Im}(A) = \mathbb{1}^\perp$.

First, we prove that $\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}$ is the orthogonal projector on $\operatorname{span} \mathbb{1}$. It is orthogonal, and since

$$\left(\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}\right)\left(\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}\right) = \frac{\mathbb{1}^\mathsf{T}\mathbb{1}}{n^2}\mathbb{1}\mathbb{1}^\mathsf{T} = \frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T},$$

it is a projection. Finally, note that x is fixed by $\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}$ if and only if it is a multiple of $\mathbb{1}$:

$$\left(\frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}\right) x = x \iff x = \left(\frac{\mathbb{1}^\mathsf{T} x}{n}\right)\mathbb{1}.$$

Therefore, Equation 2.43 ensures that $I - \frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T}$ is the orthogonal projector on the orthogonal complement of $\operatorname{span} \mathbb{1}$.

With this information, Equation (a) of the definition of Moore-Penrose pseudoinverse can be written as

$$\begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix} \begin{bmatrix} A^\dagger[\{i\}^c, \{i\}^c] & A^\dagger[\{i\}^c, i] \end{bmatrix} = \begin{bmatrix} A^\dagger[\{i\}^c, \{i\}^c] & A^\dagger[\{i\}^c, i] \\ \mathbb{1}^\mathsf{T} A^\dagger[\{i\}^c, \{i\}^c] & \mathbb{1}^\mathsf{T} A^\dagger[\{i\}^c, i] \end{bmatrix} = \begin{bmatrix} I - \frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T} & -\frac{1}{n}\mathbb{1} \\ -\frac{1}{n}\mathbb{1}^\mathsf{T} & 1 - \frac{1}{n} \end{bmatrix},$$

which implies that

$$A^\dagger[\{i\}^c, \{i\}^c] = I - \frac{1}{n}\mathbb{1}\mathbb{1}^\mathsf{T},$$
$$A^\dagger[\{i\}^c, i] = -\frac{1}{n}\mathbb{1}^\mathsf{T},$$

or, assuming the first columns are indexed by $\{i\}^c$,

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

4.2 Effective Resistances as Marginal Probabilities

Proposition 4.16. Let $G = (V, E, \psi)$ be a connected graph. Let $e_0 \in E$ be such that $|\psi(e_0)| = 2$. Denote by i and j the elements of $\psi(e_0)$. 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 $\{i\}^c$, the equality $L_G \mathbb{1} = 0$ turns into

$$\begin{bmatrix} L_G[\{i\}^c, \{i\}^c] & L_G e_i \\ -e_i^\mathsf{T} L_G & (L_G)_{ii} \end{bmatrix} \begin{bmatrix} \mathbb{1} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Solving this system for both $L_G e_i$ and $(L_G)_{ii}$, we arrive at

$$L_G = \begin{bmatrix} L_G[\{i\}^c, \{i\}^c] & -L_G[\{i\}^c, \{i\}^c]\mathbb{1} \\ -\mathbb{1}^\mathsf{T} L_G[\{i\}^c, \{i\}^c] & \mathbb{1}^\mathsf{T} L_G[\{i\}^c, \{i\}^c]\mathbb{1} \end{bmatrix} = \begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix} L_G[\{i\}^c, \{i\}^c] \begin{bmatrix} I & -\mathbb{1} \end{bmatrix}.$$

Note that since G is connected, it has at least a spanning tree, so that $\det(L_G[\{i\}^c, \{i\}^c])$ is nonzero, and, therefore, $L_G[\{i\}^c, \{i\}^c]$ has both full row and full column rank. Also, since $\begin{bmatrix} I - \mathbb{1} \end{bmatrix}$ is surjective, it has full row rank, and its transpos has full column rank. Therefore, Proposition 4.13 applied twice ensures that

$$L_G^\dagger = \begin{bmatrix} I & -\mathbb{1} \end{bmatrix}^\dagger L_G[\{i\}^c, \{i\}^c]^{-1} \begin{bmatrix} I \\ -\mathbb{1}^\mathsf{T} \end{bmatrix}^\dagger.$$

From Example 4.14 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 4.17. Let $G = (V, E, \psi)$ be a connected graph. Let $e_0 \in E$ be such that $|\psi(e_0)| = 2$. Denote by i and j the elements of $\psi(e_0)$. Then

$$\frac{\Phi(G/e_0)}{\Phi(G)} = (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}.$$

Proof. Note that

$$L_G = L_{G-e_0} + (e_i - e_j)(e_i - e_j)^\mathsf{T},$$

so that

$$L_{G-e_0}[\{i\}^c, \{i\}^c] = L_G[\{i\}^c, \{i\}^c] - e_j e_j^\mathsf{T}.$$

Applying Lemma 2.40, we then have that

$$\Phi(G - e_0) = \Phi(G)(1 - (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}).$$

However, $\Phi(G) = \Phi(G/e_0) + \Phi(G - e_0)$, and we conclude that

$$\frac{\Phi(G/e_0)}{\Phi(G)} = 1 - \frac{\Phi(G - e_0)}{\Phi(G)} = 1 - (1 - (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}) = (L_G[\{i\}^c, \{i\}^c]^{-1})_{jj}.$$

□

Both propositions lead into a way to calculate the probability of an edge to belong to the tree output of $\mathcal{A}(G)$ using the pseudoinverse instead of a determinant. This is interesting because it only demands 4 entries of the pseudoinverse matrix to be known, and our next algorithm will explore this.

Theorem 4.18. 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 $T \in \mathcal{T}_G$

$$\mathbb{P}(\mathcal{A}(G) = T) = \frac{1}{\Phi(G)} \prod_{e \in T} w(e).$$

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:

$$\mathbb{P}(e_0 \in \mathcal{A}(G)) = w(e_0)\frac{\Phi(G/e_0)}{\Phi(G)} = w(e_0)(L_G[\{i\}^c, \{i\}^c]^{-1})_{jj} = w(e_0)(e_i - e_j)^\mathsf{T} L_G^\dagger (e_i - e_j). \qquad \square$$

4.3 The Harvey and Xu algorithm

Part II

Random Walk-Based Algorithms

Chapter 5

Random Walks

5.1 Markov Chains

A *stochastic process* in a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ 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 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 a *Markov chain* if

$$\mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k, \ldots, X_0 = x_0) = \mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k) \tag{5.1}$$

for every $k \in \mathbb{N}$ and $x \colon \{0, \ldots, k+1\} \to V$ such that both conditional probabilities are well defined, i.e., such that $\mathbb{P}(X_k = x_k, \ldots, X_0 = x_0) > 0$. Condition (5.1) is sometimes called the *Markov property*.

Let X be a Markov chain. For each $k \in \mathbb{N}$, the *transition matrix at time* k is the matrix $P_k \colon V \times V \to \mathbb{R}$ defined by $(P_k)_{uv} = \mathbb{P}(X_{k+1} = v \mid X_k = u)$ for each $(u, v) \in V \times V$. If $P_k = P_\ell$ for each $k, \ell \in \mathbb{N}$, the Markov chain is *time-homogeneous* or *stationary*, and the common value $P \colon V \times V \to \mathbb{R}$ is the *transition matrix*.

5.2 Arrival time and cover time

Lets take a look at the random variables taking values at R^*. The first interesting property is the following:

Theorem 5.2 (Infimum of random variables)**.** Let $(X_k)_{k \in \mathbb{N}}$ be a sequence of random variables from the measurable space (Ω, Σ) to $\mathbb{R} \cup \{\text{-}\infty, +\infty\}$. We have that the function $Y : \omega \to \mathbb{R}^*$ defined by

$$Y(\omega) = \inf_{k \in \mathbb{N}} X_k(\omega)$$

is a random variable.

Proof. Using that it is enough to prove the measurability of a function on a generator set of the sigma algebra in the image, it is enough to demonstrate that, for any $\alpha \in \mathbb{R}$, the set $Y^{-1}([-\infty, \alpha))$ is measurable. Note that, for any such set,

$$\begin{aligned} Y^{-1}([-\infty, \alpha)) =& \{\omega \in \Omega : Y(\omega) < \alpha\} \\ =& \left\{\omega \in \Omega : \inf_{k \in \mathbb{N}} X_k(\omega), \alpha\right\} \\ =& \{\omega \in \Omega : \exists k \in \mathbb{N} : X_k(\omega) < \alpha\} \\ =& \bigcup_{k \in \mathbb{N}} X_k^{-1}([-\infty, \alpha)) \end{aligned} \tag{5.3}$$

Therefore the set is a countable union of sets that are measurable since every X_k is a random variable, and the $[-\infty, \alpha)$ are measurables. ☐

With this tool in hand, given a markov chain $(X_k)_{k\in\mathbb{N}}$, I can define the random variables $(Y_k)_{k\in\mathbb{N}}$, for any fixed natural k,

$$Y_k(\omega) = \begin{cases} k, & \text{if } X_k(\omega) = j \\ +\infty, & \text{otherwise} \end{cases} \tag{5.4}$$

where $j \in V$ is a fixed point in the state space of X_k.

It is possible to state that for any $k \in \mathbb{N}$, Y_k is measurable, by applying the result on the measurability of function with countable image, and noticing that

$$Y_k(\Omega) = \{k, +\infty\}$$

Therefore, the arrival time $A_j : \Omega \to \mathbb{R}^*$ can be defined as

$$A_j(\omega) := \inf_{k\in\mathbb{N}} Y_k(\omega)$$

which is measurable by the result proven on the begining of this section.

Chapter 6

The Aldous-Broder Algorithm