A Randomized $O(\ln n/\ln\ln n)$-Approximation Algorithm for the Metric Asymmetric Traveling Salesman Problem

Germano Hüning Neuenfeld
Supervisor: Marcel K. de Carli Silva

March 10, 2021

Abstract

The Traveling Salesman Problem (TSP) is central in theoretical computer science and has a myriad of applications. A generalization of the TSP, the Asymmetric Traveling Salesman Problem (ATSP), has a metric version, denoted by mATSP, which admits an approximation algorithm with a polynomial-time computable approximation ratio. This monograph presents a randomized approximation algorithm for the mATSP due to Asadpour, Goemans, Madry, Oveis Gharan, and Saberi [2], which represented a breakthrough for this problem. Moreover, to understand the Asadpour *et al.* algorithm, this work presents a collection of tools from areas such as combinatorial optimization, linear programming, and probability theory whose use is standard in the study of approximation algorithms.

Keywords: asymmetric traveling salesman problem, thin trees, concentration bounds, random spanning trees, randomized rounding, negatively correlated random variables.

Resumo

O Problema do Caixeiro Viajante (TSP) é central em teoria da computação e tem uma infinidade de aplicações. Uma generalização do TSP, o Problema do Caixeiro Viajante Assimétrico (ATSP), tem uma versão métrica, denotada por mATSP, que admite um algoritmo de aproximação com uma razão de aproximação que é computável em tempo polynomial. Esta monografia apresenta um algoritmo de aproximação aleatorizado para o mATSP devido a Asadpour, Goemans, Madry, Oveis Gharan e Saberi [2] que representou um grande avanço para esse problema. Além disso, para compreender o algoritmo devido a Asadpour *et al.*, este trabalho apresenta uma coleção de ferramentas de áreas como otimização combinatória, programação linear, e teoria da probabilidade cujo uso é padrão no estudo de algoritmos de aproximação.

Palavras-chave: problema do caixeiro viajante assimétrico, árvores finas, limitantes de concentração, árvores geradoras aleatórias, arredondamento aleatorizado, variáveis aleatórias negativamente correlacionadas.

Acknowledgements

First, thanks to my family: my father Lairton Neuenfeld, my mother Ilisete Luiza Hüning, and my sister Julia Hüning Neuenfeld. They offer me the patience, support, and understanding I needed to keep going. Also, I want to thank my friends for the support and advice, in particular to the long hours of talking with Victor Aliende da Matta and Thiago Estrela Montenegro.

I am immensely grateful for having Professor Marcel Kenji de Carli Silva as my supervisor. His mentoring helped me to develop this work to way more than I initially could. At the same time, his mentoring showed me that much more can be done. I am also thankful to Professor Nina Hirata for her kindness and patience during the development of the monograph.

Finally, I want to dedicate this work to Daisi Oreques da Silva (in memoriam), the “tia Dai”, who left us last year. Your smile will always be in my mind.

Contents

Chapter 1

Introduction

The Traveling Salesman Problem (TSP) is central in theoretical computer science and has a myriad of applications (see Cook [7, Chapter 3]) that range from those that provide the name of the problem, namely designing minimum-cost road trips, to even surprising ones such as mapping genomes. The usual presentation of the problem is with a road trip through cities. Given a set of cities and the distance between each pair, we want to know how one can visit all cities exactly once and come back to where the journey began while traveling the smallest distance possible. This is a formulation of the Traveling Salesman Problem (TSP). The Asymmetric Traveling Salesman Problem (ATSP) is a generalization of the TSP where the distances between the cities can be different according to the direction of movement (hence the "asymmetric").

The TSP and the ATSP are NP-hard; so both problems cannot be solved in polynomial time unless P = NP. Moreover, neither admits an approximation algorithm with a polynomial-time computable approximation ratio unless P = NP (see Vazirani [20, Theorem 3.6]). However, one can impose some assumptions that allow these problems to be approximated while still being interesting from both an application and a theoretical perspective. The Metric (Asymmetric) Traveling Salesman Problem, denoted by mTSP (mATSP), is the TSP (ATSP) when one imposes that the distances between the cities satisfy the triangular inequality. In this monograph, we present a randomized approximation algorithm for the mATSP due to Asadpour, Goemans, Madry, Oveis Gharan, and Saberi [2].

The Asadpour *et al.* algorithm is a mark in the history of the mATSP. First, when it was published in 2010 as a $o(\log n)$-approximation algorithm for the mATSP, it represented the break of a barrier of over 25 years since the $\Theta(\log n)$-approximation algorithm due to Frieze, Galbiati, and Maffioli [9]. Second, it represents an application of a collection of areas and their tools that is standard in the study of approximation algorithms. Starting from the classic area of combinatorial optimization with problems involving minimum cuts and minimum-cost circulations, it goes through linear programming with the Held-Karp relaxation of ATSP, the ellipsoid method, and the equivalence of separation and optimization problems, and arrives at a set of probability tools such as Chernoff bounds and the sampling of spanning trees using a maximum entropy distribution. Also, the algorithm uses the recent definition of thin trees, which has been shown to be very fruitful; for instance, Anari and Gharan [1] give a bound on the integrality gap of the Held-Karp relaxation of ATSP using this definition.

Both the importance of the mATSP and the tools used to design this algorithm would already make the study of this algorithm worthwhile. However, this algorithm also has a beautiful and remarking resemblance with the algorithm due to Christofides [6] which is a 3/2-approximation algorithm for the mTSP. First, each one finds a specific spanning tree. While this is a fairly easy task in Christofides's algorithm, in the Asadpour *et al.* algorithm, it is the most difficult and central part of the algorithm. Then Christofides's algorithm builds an Eulerian graph using an optimum solution of the minimum-cost perfect matching problem, while the Asadpour *et al.* algorithm builds an Eulerian digraph using an optimum solution of the minimum-cost integer circulation problem. Finally, both algorithms use a idea of shortcutting the Eulerian graph/digraph into the final result, a Hamiltonian circuit of minimum cost.

In this monograph, we dive into this interesting connection of subjects and study the algorithm of Asadpour *et al.* In Chapter 2, we present a few notations, definitions, and results used throughout the text, such as those involving linear programming and graph theory. In Chapter 3, we present the ATSP (and its metric version mATSP) and the Asadpour *et al.* algorithm; this algorithm has three steps that we prove we can perform in polynomial time in the subsequent chapters. In Chapter 4, we prove the Max-Flow Min-Cut theorem, and we present the Edmonds-Karp Algorithm to find a minimum cut in polynomial time; also, we present Hoffman's circulation theorem. In Chapter 5, we present the probability tools needed for the randomized part of the algorithm, and we show how we can perform the second step of the Asadpour *et al.* algorithm in polynomial time. In particular, we present randomized swap rounding (RSR), an algorithm we use for sampling random spanning trees. This sampling is done in [2] using a maximum entropy distribution; however, RSR provides a simpler approach for such sampling with stronger guarantees. Finally, in Chapter 6, we show the method of ellipsoids, and we present the theorem of equivalence of separation and optimization problems; this last result is used to show the first step of the algorithm can be done in polynomial time.

Finally, we should also remark on the still high activity in the area. In 2018, Svensson, Tarnawski, and Végh [18] provided a *constant-factor* approximation algorithm for the mATSP and gave a constant upper bound on the integrality gap of the Held-Karp relaxation of ATSP. Also, in an even more recent development in 2019, Traub and Vygen [19] improved these constants.

Chapter 2

Preliminaries

This chapter presents some basic notation, definitions and results that will pervade the text. Throughout the text, the formal definitions will be given by **bold** words, while when we have an informal definition, or when we just want to give emphasis to the concept we will use *italic* words.

2.1 Notation

The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are the sets of natural, integer, rational, and real numbers, respectively. Define $[n] \coloneqq \{1, \dots, n\}$ for each $n \in \mathbb{N}$. If $\mathbb{I} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$, define

$$\mathbb{I}_+ \coloneqq \{\, \alpha \in \mathbb{I} : \alpha \geq 0 \} \quad \text{and} \quad \mathbb{I}_{++} \coloneqq \{\, \alpha \in \mathbb{I} : \alpha > 0 \}\,. \tag{2.1}$$

We will use Minkowski notation to sum sets, that is, if U and V are sets, then $U+V \coloneqq \{\, u+v : u \in U, v \in V \}$. We abbreviate $v + U \coloneqq \{v\} + U$ for a singleton $\{v\}$ and a set U.

The **Iverson bracket** of a predicate P is defined by

$$[P] \coloneqq \begin{cases} 1\,, & \text{if } P \text{ is true}\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{2.2}$$

2.2 General Math

We start with some set theory, relations, and functions. We presuppose the reader has some familiarity with these concepts. Even so, we present some notation and definitions that we judge are worth stating to avoid confusions or misinterpretations.

Let A be a finite set, and let k be a nonnegative integer with $k \leq |A|$. Then a ***k*-set** of A is a subset of A with k elements. Moreover, we denote by $\binom{A}{k}$ the set of all k-sets of A.

Let A, B be sets. We denote by B^A the set of all functions from A to B, and if $A = [n]$ for some $n \in \mathbb{N}$, we abbreviate $B^{[n]}$ by B^n. Let $a \in A$, and let $b \in B$. We abbreviate $ab \coloneqq (a, b)$, and we define $ab^{-1} \coloneqq ba$. These two definitions regarding ordered pairs are useful in many contexts; for instance, to write the entries of a matrix, to write an arc of a digraph, and to define the reverse digraph of a digraph.

Let $f \colon A \to B$ be a function, and let $a \in A$. The **image** or **value** of a under f is $f(a)$. Let X and Y be subsets of A and B, respectively. The **image** of X under f is the set

$$f[X] \coloneqq \{\, b \in B : \text{there is } x \in X \text{ such that } f(x) = b \}\,. \tag{2.3}$$

The **preimage** or **inverse image** of Y under f is the set

$$f^{-1}[Y] \coloneqq \{\, a \in A : f(a) \in Y \}\,; \tag{2.4}$$

we abbreviate $f^{-1}[y] \coloneqq f^{-1}[\{y\}]$ for each $y \in B$. Usually, one introduces the definitions of image of a subset of the domain, and of preimage of a subset of the codomain, under a function using parentheses instead of square brackets. However, this may lead to confusion since the domain and/or codomain of the function can have a set as an element. For instance, if $A = \{1, 2, \{1, 2\}\}$, then $f(\{1, 2\})$ refers to the value of $\{1, 2\}$ under f, while $f[\{1, 2\}]$ refers to the image of set $\{1, 2\}$ under f.

Let $f\colon A \to B$ be a function. Then f is invertible if and only if f is injective and surjective. One can show that if f is an invertible function, then it has a unique inverse that we denote by f^{-1}. Since f^{-1} is a function, the above definitions of *value*, *image*, and *preimage* also apply to f^{-1}. Note that if f is invertible, then $f^{-1}[Y]$, for a subset Y of B, can mean the preimage of Y under f or the image of Y under f^{-1}. However, in this case, both interpretations provide the same set; so there is no ambiguity in defining both concepts with the same notation. Even so, one must keep in mind that the expression $f^{-1}[Y]$, for a subset Y of B, does not mean that the function f has an inverse function f^{-1}.

Proposition 2.1. Let $f\colon A \to B$ be an injective function. Then

(i) $f[X \cap Y] = f[X] \cap f[Y]$ for each $X, Y \subseteq A$, and

(ii) $f^{-1}[f[X]] = X$ for each $X \subseteq A$.

Let $f\colon A \to \mathbb{R}$ be a function. The **support** of f is the set $\operatorname{supp}(f) \coloneqq \{\, a \in A : f(a) \neq 0 \}$. Let $g\colon B \to C$ be a function, and let S be a subset of B. The **restriction** of g to S is the function $g{\restriction_S}\colon s \in S \mapsto g(s)$.

A relation $\leq$ on a set A is a **partial order** on A if, for each $a, b, c \in A$,

(i) (Reflexivity) $a \leq a$,

(ii) (Antisymmetry) if $a \leq b$ and $b \leq a$, then $a = b$, and

(iii) (Transitivity) if $a \leq b$ and $b \leq c$, then $a \leq c$.

A **partially ordered set**, or **poset**, is an ordered pair $(P, \leq)$, where P is a set and $\leq$ is a partial order on P. Let $(P, \leq)$ be a poset, and let S be a subset of P. An element a of S is **maximal** in S if $\{\, b \in S : a \leq b \} = \{a\}$, i.e., there is no $b \in S$ such that $a \leq b$ and $a \neq b$ (in words: if we translate $c \leq d$ to d *greater than or equal to* c for each $c, d \in S$, then a is maximal in S if there is no element b in S, distinct of a, that is greater than or equal to a). An element a of S is **maximum** in S if $b \leq a$ for each $b \in S$. Note that if S has a maximum, it is unique. Also, note that the maximum element in S, if it exists, is a maximal element, but the converse is not necessarily true.

A relation $<$ on a set A is a **strict total order** on A if, for each $a, b, c \in A$,

(i) (Asymmetry) if $a < b$, then $b \not< a$,

(ii) (Transitivity) if $a < b$ and $b < c$, then $a < c$, and

(iii) (Semiconnexity) $a < b$ or $b < a$ or $a = b$.

Also, one can show that a binary relation R on a set A is asymmetric and semiconnex if and only if it is trichotomous, that is, for each $a, b \in A$, exactly one of aRb, bRa, and $a = b$ holds. Thus, a strict total order can be equivalently defined as a relation that satisfies transitivity and trichotomy.

Now we turn into some definitions involving sequences and series. They concern the proof of Proposition 2.3, an important inequality involving the exponential function that appears in some proofs.

A **sequence** (of real numbers) is a function from $\mathbb{N}$ to $\mathbb{R}$. We have a special notation for a function of this form. If a is a sequence, for each $n \in \mathbb{N}$, we denote $a(n)$ by a_n, and we say a_n is the n-th **term** of the sequence; also, the values of a, i.e, $a_1, a_2, a_3, \ldots$, are the **terms** of the sequence a. We write $(a_n)_{n\in\mathbb{N}}$, or simply (a_n), to indicate a sequence a. Sometimes, we may include zero and consider $(a_n)_{n\,\in\mathbb{N}\cup\{0\}}$ a sequence.

A sequence (a_n) **converges** to a real number a if

$$\text{for each } \varepsilon > 0, \text{ there exists } n_0 \in \mathbb{N} \text{ such that for each } n \in \mathbb{N}, \text{ if } n > n_0, \text{ then } |a_n - a| < \varepsilon. \tag{2.5}$$

If a sequence (a_n) converges to a real number a, we say (a_n) is **convergent**; in this case, we say a is limit of (a_n), and we write $\lim a_n = a$. Actually, one can show that if a sequence is convergent, then its limit is unique, and so we can say *the* limit of a sequence if the sequence is convergent.

Now we present an important family of sets for the definition of the limit of a sequence, the open intervals. Let $a \in \mathbb{R}$, and let $\varepsilon > 0$. The open interval $(a-\varepsilon, a+\varepsilon)$ is the set $\{ x \in \mathbb{R} : a - \varepsilon < x < a + \varepsilon \}$. Then a real $x \in \mathbb{R}$ is in $(a-\varepsilon, a+\varepsilon)$ if and only if $|x - a| < \varepsilon$. Moreover, $(a-\varepsilon, a+\varepsilon)$ is called the **ε-neighborhood** of a. Thus, we can provide a characterization of the limit of a sequence in terms of *neighborhoods* of the limit in the real line.

A sequence (a_n) converges to a real number a if and only if

$$\text{for each } \varepsilon > 0, \text{ there exists } n_0 \in \mathbb{N} \text{ such that for each } n \in \mathbb{N}, \text{ if } n > n_0, \text{ then } a_n \in (a-\varepsilon, a+\varepsilon). \tag{2.6}$$

Let (a_n) be a sequence. Define the corresponding sequence (s_n) as

$$s_n = \sum_{i=1}^{n} a_i \qquad \text{for each } n \in \mathbb{N}\,.$$

Denote the sequence (s_n) by

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots\,. \tag{2.7}$$

We say the expression $\sum_{n=1}^{\infty} a_n$ is a **(infinite) series**. The terms s_n are the **partial sums** of the series $\sum_{n=1}^{\infty} a_n$. Moreover, we say the series $\sum_{n=1}^{\infty} a_n$ converges to s if the sequence (s_n) converges to s; in this case, we write $\sum_{i=1}^{\infty} a_n = s$, and we say s is the *sum* of the series.

Note that when we write $\sum_{n=1}^{\infty} a_n = s$, we mean that s is the limit of a sequence of sums (in case, (s_n)), and not that s is the "addition" of infinite terms of s_n. It may be convenient to consider sequences $(a_n)_{n \in \mathbb{N} \cup \{0\}}$, and then we may consider $\sum_{n=0}^{\infty} a_n$ a series. We write $\sum a_n$ to denote $\sum_{n=1}^{\infty} a_n$ or $\sum_{n=0}^{\infty} a_n$, when possible.

A series of the form $\sum_{n=0}^{\infty} a_n x^n$, for a sequence (a_n) and a scalar $x \in \mathbb{R}$, is a **power series**.

Lemma 2.2. Let (a_n) be a convergent sequence such that $a_n \geq 0$ for each $n \in \mathbb{N}$. Then $\lim a_n \geq 0$.

Proof. The proof is by contradiction. Suppose $\lim a_n = a < 0$, and set $\varepsilon := |a| > 0$. Then, by the characterization of the limit of a sequence (see (2.6)), there exists $n_0 \in \mathbb{N}$ such that for each $n \in \mathbb{N}$, if $n > n_0$, then $a_n \in (a-\varepsilon, a+\varepsilon)$. Therefore, we have a contradiction with $a_{n_0+1} < a + \varepsilon < 0$, for instance. □

Proposition 2.3. For each $x \in \mathbb{R}$,

$$e^x \geq 1 + x\,.$$

Proof. We divide the proof into three cases. For the first case, suppose $x \leq -1$. Then $e^x > 0$ while $1 + x \leq 0$.

Now by definition of the exponential function e^x as a power series, we have, for each $x \in \mathbb{R}$,

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} = 1 + x + \sum_{i=2}^{\infty} \frac{x^i}{i!} \tag{2.8}$$

Suppose $x \geq 0$. Then each term of the sum $\sum_{i=2}^{\infty} \frac{x^i}{i!}$ is nonnegative, and so $e^x \geq 1 + x$ by the RHS of (2.8).

Finally, suppose $-1 < x < 0$. Set (s_n) to be the sequence such that $s_n = \sum_{i=2}^{n+1} \frac{x^i}{i!}$ for each $n \in \mathbb{N}$. We claim $s_n \geq 0$ for each $n \in \mathbb{N}$. Indeed, (s_n) is an alternating sequence, that is, the sign of the sequence terms changes between consecutive terms; the first term of s_n is positive; moreover, as n increases, s_n decreases in modulus. Thus, by Lemma 2.2, $\lim s_n \geq 0$, i.e., $\sum_{i=2}^{\infty} \frac{x^i}{i!} \geq 0$. Therefore, by the RHS of (2.8), $e^x \geq 1 + x$. □

2.3 Graph Theory

A **graph** is an ordered triple (V, E, ψ), where V and E are sets, and ψ is a function from E to $\binom{V}{2} \cup \binom{V}{1}$. If $G = (V, E, \psi)$ is a graph, the elements of V and E are the **vertices** and the **edges** of G, respectively, and ψ is called the **incidence function** of G. If G is a graph, we denote by $V(G)$ the set of vertices of G, by $E(G)$

the set of edges of G, and by ψ_G the incidence function of G. Thus, we may refer to the vertex set, the edge set, and the incidence function of a graph even though we have not given a specific symbol for each of them.

Let $G = (V, E, \psi)$ be a graph. An edge $e \in E$ is a **loop** if $|\psi(e)| = 1$. Two distinct edges $e, f \in E$ that are not loops and for which $\psi(e) = \psi(f)$ are called **parallel edges**. The graph G is **simple** if G has no loops nor parallel edges. Equivalently, G is simple if ψ is a function from E to $\binom{V}{2}$ and ψ is injective. When ψ is the identity function $e \in E \mapsto e$ (and so $E \subseteq \binom{V}{2} \cup \binom{V}{1}$), instead of writing $G = (V, E, \psi)$, we may say that G is the pair (V, E); in other words, if we say that $H = (U, F)$ is a graph, it is implicit that $F \subseteq \binom{U}{2} \cup \binom{U}{1}$ and that ψ_H is the identity function.

Let $G = (V, E, \psi)$ be a graph. We abbreviate $uv \coloneqq \{u, v\}$ for any $u, v \in V$. If $u, v \in V$ and $e \in E$ with $\psi(e) = uv$, we say e **joins** u and v, u and v are called the **ends** of e, and e is **incident** to u and v. Two vertices $u, v \in V$ are called **adjacent**, and u is a **neighbor** of v (and v is a neighbor of u) if $\psi(e) = uv$ for some $e \in E$. The **neighborhood** of a vertex $v \in V$, denoted by $N_G(v)$, is the set of neighbors of v. If $U \subseteq V$, we say

$$E_\psi[U] \coloneqq \{\, e \in E : \psi(e) \subseteq U \,\}$$

is the edge set of G **induced** by U, i.e., $E_\psi[U]$ is the set of edges of G with both ends in U. When the incidence function ψ is clear from context, we may use $E[U] \coloneqq E_\psi[U]$. Also, we abbreviate $E_\psi[v] \coloneqq E_\psi[\{v\}]$ for each $v \in V$. If $U \subseteq V$, we define

$$\delta_G(U) \coloneqq \{\, e \in E : \psi(e) = uv \text{ for some } u \in U \text{ and } v \notin U \,\},$$

i.e., $\delta_G(U)$ is the set of edges of G with exactly one end in U. When the graph G is clear from context, we may use $\delta(U) \coloneqq \delta_G(U)$ for any $U \subseteq V$. Moreover, we abbreviate $\delta_G(v) \coloneqq \delta_G(\{v\})$ for each $v \in V$.

A graph G is **complete** if G is simple and any two distinct vertices are adjacent.

Let $G = (V, E, \psi)$ be a graph. A **walk** in G is a sequence $(v_0, e_1, v_1, \ldots, e_k, v_k)$ for some $k \geq 0$, where $v_0, v_1, \ldots, v_k$ are vertices of G and, for each $i \in [k]$, e_i is an edge of G with ends v_{i-1} and v_i. Let $W = (v_0, e_1, v_1, \ldots, e_k, v_k)$ be a walk in G for some $k \geq 0$. The walk W is said to **connect** v_0 and v_k, and to **traverse** vertices $v_0, \ldots, v_k$ and edges $e_1, \ldots, e_k$; also, we say W is a walk **from** v_0 **to** v_k (or **between** v_0 and v_k). Moreover, we say W is a v_0-v_k **walk**, and for $S, T \subseteq V$, the walk W is called an S-T walk if $v_0 \in S$ and $v_k \in T$, i.e., if W is a walk from a vertex of S to a vertex of T. The vertex v_0 is the **starting vertex** or the **first vertex** of W while the vertex v_k is the **end vertex** or the **last vertex** of W. Vertices v_0 and v_k are the **end vertices** of W. The **length** of W is the integer k, i.e., the number of edges of W. A walk is **odd** (**even**, resp.) if its length is odd (even, resp.). If $P = (v_0, e_1, v_1, \ldots, e_k, v_k)$ and $Q = (u_0, f_1, u_1, \ldots, f_\ell, u_\ell)$ are walks such that $u_0 = v_k$, the **concatenation** of P and Q is the walk $P \cdot Q \coloneqq (v_0, e_1, v_1, \ldots, e_k, v_k, f_1, u_1, \ldots, f_\ell, u_\ell)$. If $W = (v_0, e_1, v_1, \ldots, e_k, v_k)$ is a walk in G and $i < j$ are integers in $\{0, 1, \ldots, k\}$, then the subsequence $(v_i, e_i, v_{i+1}, \ldots, e_j, v_j)$ of W, denoted by $v_i W v_j$, is called the **segment** of W from v_i to v_j. We write $V(W)$ and $E(W)$ for the set of vertices traversed by W and the set of edges traversed by W, respectively. Thus, we may consider W as the subgraph $(V(W), E(W), \psi\restriction_{E(W)})$ of G. If G is simple, then W is determined by the sequence of vertices $(v_0, \ldots, v_k)$, and we may identify W with such sequence.

A **trail** in a graph G is a walk in G with no repeated edges. A **path** in a graph G is a walk in G with no repeated vertices. A walk in a graph G is **closed** if it has positive length and equal end vertices. A **cycle** in a graph G is a closed trail in G. A **circuit** in a graph G is a cycle $(v_0, e_1, v_1, \ldots, e_k, v_k)$ in G for some $k \geq 1$, where $v_1, \ldots, v_k$ are all distinct. A trail P in a graph G is called **Eulerian** if each edge of G is traversed exactly once by P. A graph G is called **Eulerian** if it has an Eulerian cycle.

Let $G = (V, E, \psi)$ be a graph. A graph H is a **subgraph** of G if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, and $\psi_H = \psi_G\restriction_{E(H)}$. If H is a subgraph of G, we say H is **contained** in G or G **contains** H, and we write $H \subseteq G$ or $G \supseteq H$, respectively. In addition, if $H \subseteq G$, we also say G is a **supergraph** of H. Note that G is a subgraph and a supergraph of itself. Thus, if $H \subseteq G$ and $H \neq G$, we say H is a **proper subgraph** of G and G is a **proper supergraph** of H. If $H \subseteq G$ and $V(H) = V(G)$, then H is a **spanning subgraph** of G. The subgraph of G **induced** by a nonempty subset U of V is the graph $G[U] \coloneqq (U, E[U], \psi\restriction_{E[U]})$.

Let $G = (V, E, \psi)$ be a graph. Let U and F be subsets of V and E, respectively. Moreover, let H, H' be subgraphs of G. Then

$$
\begin{aligned}
G - U &\coloneqq G[V \setminus U]\,, \\
G - F &\coloneqq (V, E \setminus F, \psi\restriction_{E\setminus F})\,, \\
H + U &\coloneqq (V(H) \cup U, E(H), \psi\restriction_{E(H)})\,, \\
H + F &\coloneqq (V(H), E(H) \cup F, \psi\restriction_{E(H)\cup F})\,, \\
H + H' &\coloneqq (V(H) \cup V(H'), E(H) \cup E(H'), \psi\restriction_{E(H)\cup E(H')})\,.
\end{aligned}
\tag{2.9}
$$

To simplify, if x is a vertex or an edge of G, we write $G - x$ and $H + x$ instead of $G - \{x\}$ and $H + \{x\}$, respectively.

A graph G is **connected** if for each $u, v \in V(G)$ there exists a walk in G connecting u and v. Let G be graph. A subgraph C of G is a **(connected) component** of G if C is connected and any subgraph H of G which is a proper supergraph of C is not connected (**disconnected**). Equivalently, a subgraph C of G is a component of G if C is a maximal connected subgraph of G, that is, for the poset $(S, \subseteq)$, where S is the set of subgraphs of G, the subgraph C is a maximal element of the subset of connected graphs of S. If C is a component of G, we may identify the graph C with the vertex set $V(C)$; thus, we may write, for instance, $C \subseteq V(G)$.

Let G be a graph. The graph G is **acyclic** if it has no circuits (or, equivalently, if it has no cycles); an acyclic graph is also called a **forest**. A **tree** is a connected forest. A **spanning tree** of G is a spanning subgraph of G that is a tree.

A **digraph** is an ordered triple (V, A, ψ) defined similarly to a graph, except that elements of A are called **arcs** and the incidence function is $\psi : A \to V \times V$. Let $D = (V, A, \psi)$ be a digraph. An arc $a \in A$ is a **loop** if $\psi(a) = (v, v)$ for some $v \in V$. Two distinct arcs $a, b \in A$ that are not loops and for which $\psi(a) = \psi(b)$ are called **parallel arcs**. The digraph D is **simple** if D has no loops nor parallel arcs. Similar to when we write a graph as a pair, when ψ is the identity function $a \in A \mapsto a$ (and so $A \subseteq V \times V$), we may omit the incidence function and say that D is the pair (V, A).

From Section 2.2, recall that we may write an arbitrary ordered pair (i, j) as ij; when dealing with graphs and digraphs, we use such convention in a way that, from the context, one can correctly identify whether ij means an arc (i, j) or an edge $\{i, j\}$.

Let $D = (V, A, \psi)$ be a digraph. If $\psi(a) = uv$ for some arc $a \in A$, then u and v are the **ends** of a, and u is called the **tail** of a while v is called the **head** of a; in addition, we say a is **incident** to u and v, the vertices u and v are called **adjacent**, and we say a **leaves** u and **enters** v. Similarly, for $U \subseteq V$, if $\psi(a) = uv$ for an arc $a \in A$ and vertices $u \in U, v \notin U$, then we say *a leaves U* and *enters $\overline{U} \coloneqq V \setminus U$*. If $U \subseteq V$, we say

$$A_\psi[U] \coloneqq \{\, a \in A : \psi(a) \in U \times U \,\}$$

is the arc set of D **induced** by U, i.e., $A_\psi[U]$ is the set of arcs of A with both ends in U. When the incidence function ψ is clear from context, we may use $A[U] \coloneqq A_\psi[U]$. Also, we abbreviate $A_\psi[v] \coloneqq A_\psi[\{v\}]$ for each $v \in V$. If $U \subseteq V$, we define

$$
\begin{aligned}
\delta_D^{\mathrm{in}}(U) &\coloneqq \{\, a \in A : \psi(a) = uv \text{ for some } u \in \overline{U} \text{ and } v \in U \,\}\,, \text{ and} \\
\delta_D^{\mathrm{out}}(U) &\coloneqq \{\, a \in A : \psi(a) = uv \text{ for some } u \in U \text{ and } v \in \overline{U} \,\}\,,
\end{aligned}
$$

i.e., $\delta_D^{\mathrm{in}}(U)$ is the set of arcs of A that enter U while $\delta_D^{\mathrm{out}}(U)$ is the set of arcs of A that leave U. Similarly to graphs, when the digraph D is clear from context, we may use $\delta^{\mathrm{in}}(U) \coloneqq \delta_D^{\mathrm{in}}(U)$ and $\delta^{\mathrm{out}}(U) \coloneqq \delta_D^{\mathrm{out}}(U)$ for any $U \subseteq V$. Moreover, we abbreviate $\delta_D^{\mathrm{in}}(v) \coloneqq \delta_D^{\mathrm{in}}(\{v\})$ and $\delta_D^{\mathrm{out}}(v) \coloneqq \delta_D^{\mathrm{out}}(\{v\})$ for each $v \in V$.

A digraph $D = (V, A, \psi)$ is **complete** if D is simple and $\psi(A)$ is all ordered pairs (u, v) with $u, v \in V$ and $u \neq v$.

We can also define walk, trail, path, cycle, and circuit for a digraph by just replacing edges by arcs in the respective definitions for a graph. For instance, a **walk** in a digraph D is a sequence $(v_0, a_1, v_1, \ldots, a_k, v_k)$ for some $k \geq 0$, where $v_0, v_1, \ldots, v_k$ are vertices of D and, for each $i \in [k]$, a_i is an arc of D with ends v_{i-1} and v_i. Also, a **circuit** in a digraph D is a cycle $(v_0, a_1, v_1, \ldots, a_k, v_k)$ in D for some $k \geq 1$, where $v_1, \ldots, v_k$ are all distinct. A trail P in a digraph D is **Eulerian** if each arc of D is traversed exactly once by P. A digraph D is called **Eulerian** if it has an Eulerian cycle. A **Hamiltonian circuit** C in a digraph D is a circuit in D with $V(C) = V(D)$. A digraph D is called **Hamiltonian** if it has a Hamiltonian circuit.

Let $D = (V, A, \psi)$ be a digraph, and let $\pi\colon V \times V \to \binom{V}{1} \cup \binom{V}{2}$ be a function defined by $\pi(i,j) \coloneqq \{i, j\}$ for each $i, j \in V$. The **underlying graph** of D is the graph $G = (V, A, \pi\psi)$, i.e., the graph obtained when one "ignores" the orientation of the arcs; on the other hand, D is called an **orientation** of G.

A digraph $D = (V, A, \psi)$ is **strongly connected** if for each $u, v \in V$ there exists a walk between u and v in D. A digraph is **(weakly) connected** when its underlying graph is connected.

The **reverse digraph** of D is the digraph $D^{-1} \coloneqq (V, A, \psi^{-1})$ where $\psi^{-1}(a) \coloneqq \psi(a)^{-1}$ for each $a \in A$.

2.4 Polyhedra and Linear Programming

Let V be a finite set. An **(affine) hyperplane** in $\mathbb{R}^V$ is a set of the form $\{\, x \in \mathbb{R}^V : a^\mathsf{T} x = \beta\}$ for some vector $a \in \mathbb{R}^V \setminus \{0\}$ and scalar $\beta \in \mathbb{R}$; in addition, we can call this set a **linear hyperplane** if $\beta = 0$. A **(affine) halfspace** $\mathbb{R}^V$ is a set of the form $\{\, x \in \mathbb{R}^V : a^\mathsf{T} x \leq \beta\}$ for some vector $a \in \mathbb{R}^V \setminus \{0\}$ and $\beta \in \mathbb{R}$; and similarly to hyperplanes, we can call this set a **linear halfspace** if $\beta = 0$. Usually, the term *affine* is used when the hyperplane or halfspace is not *linear*, but, as we define, affine not necessarily mean that. A **polyhedron** in $\mathbb{R}^V$ is a set of the form $\{\, x \in \mathbb{R}^V : Ax \leq b\}$ for some matrix $A \in \mathbb{R}^{U \times V}$ and vector $b \in \mathbb{R}^U$, where U is some finite set. Let U be a finite set. If $A \in \mathbb{R}^{U \times V}$ is a matrix and $b \in \mathbb{R}^U$ a vector we say $Ax \leq b$ is a **system of (linear) inequalities** and $Ax = b$ is a **system of (linear) equalities**.

Note that every hyperplane and halfspace is a polyhedron. On the other hand, any polyhedron is the intersection of a finite number of halfspaces. Moreover, a polyhedron is the solution set of a system of inequalities. Thus, if $P = \{\, x \in \mathbb{R}^V : Ax \leq b\}$ is a polyhedron in $\mathbb{R}^V$, we say the system of inequalities $Ax \leq b$ **determines** the polyhedron P; also, for a vector $d \in \mathbb{R}^V$ and a scalar $\alpha \in \mathbb{R}$, an inequality $d^\mathsf{T} x \leq \alpha$ is **valid** for P if $d^\mathsf{T} x \leq \alpha$ for each $x \in P$.

Let U, V be finite sets. Let $A \in \mathbb{R}^{U \times V}$, let $b \in \mathbb{R}^U$, and let $c \in \mathbb{R}^V$. A problem of the form

$$\text{Max } c^\mathsf{T} x \qquad \text{or} \qquad \text{Min } c^\mathsf{T} x \tag{2.10a}$$

$$\text{s.t. } Ax \leq b \qquad\qquad \text{s.t. } Ax \leq b \tag{2.10b}$$

is called a **linear program**, or for short, LP. In other words, an LP is a problem of *maximization* or *minimization* of a linear function on a vector such as c where the domain points lie in a polyhedron such as $P \coloneqq \{\, x \in \mathbb{R}^V : Ax \leq b\}$. We look at (2.10) formulation for the following general definitions of LPs. The vector c is called the **cost vector**. The linear function $x \to c^\mathsf{T} x$ is called the **objective function** or the **cost function**, and for a $x \in P$, the real number $c^\mathsf{T} x$ is called the **objective value** or **cost** of x. A point in the polyhedron P is called a **feasible solution**. The polyhedron P is called the **feasible region**. If the feasible region is nonempty, the LP is called **feasible**; otherwise, **infeasible**. If there is a point x^* in P such that $c^\mathsf{T} x^* \leq c^\mathsf{T} x$ for each $x \in P$ (minimization LP) or $c^\mathsf{T} x^* \geq c^\mathsf{T} x$ for each $x \in P$ (maximization LP), then we say the respective LP is **bounded**, the point x^* is an **optimum solution**, and $c^\mathsf{T} x^*$ is the **optimum value**; otherwise, the LP is called **unbounded**. Notice that a maximization LP can easily be turned into a minimization one or vice-versa by only inverting the sign of the cost vector, i.e., $\max\{c^\mathsf{T} x\colon Ax \leq b\} = \min\{-c^\mathsf{T} x\colon Ax \leq b\}$. So from now on we state the definitions and results on a maximization LP.

Theorem 2.4. Let V be a finite set. Let $P \subseteq \mathbb{R}^V$ be a polyhedron and $c \in \mathbb{R}^V$. Suppose P is nonempty. Then the LP

$$\begin{aligned} &\text{Max } c^\mathsf{T} x \\ &\text{s.t. } x \in P \end{aligned}$$

is either unbounded or it has an optimum solution.

2.5 Linear Algebra

We state the following results without proof. They will be used when we study ellipsoids and the ellipsoid method in Chapter 6.

Proposition 2.5 (Sherman-Morrison formula)**.** Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix and let $u, v \in \mathbb{R}^n$ be vectors. Then $A + uv^\mathsf{T}$ is invertible if and only if $1 + v^\mathsf{T} A^{-1} u \neq 0$. Moreover, if $1 + v^\mathsf{T} A^{-1} u \neq 0$, the inverse of $A + uv^\mathsf{T}$ is

$$(A + uv^\mathsf{T})^{-1} = A^{-1} - \frac{A^{-1} u v^\mathsf{T} A^{-1}}{1 + v^\mathsf{T} A^{-1} u}\,. \tag{2.11}$$

The following theorem group a collection of important results regarding positive matrices.

Theorem 2.6. Let $A \in \mathbb{R}^{V\times V}$ be a positive definite matrix. Then:

(i) There exists a unique positive definite matrix, denoted by $A^{1/2}$, that is square root of A.
(Every positive definite matrix has a unique square root that is positive definite.)

(ii) The matrix A is invertible and its inverse A^{-1} is also positive definite.
(Every positive definite matrix is invertible, and its inverse is also positive definite.)

(iii) There exists an invertible matrix $B \in \mathbb{R}^{V\times V}$ such that $A = BB^\mathsf{T} = B^\mathsf{T} B$.
(Every positive definite matrix is a product of an invertible matrix and its transpose.)

The following result gives us a way of creating positive definite matrices from invertible matrices.

Proposition 2.7. Let $A \in \mathbb{R}^{V\times V}$. If A is invertible then AA^T and $A^\mathsf{T} A$ are positive definite.

Chapter 3

The ATSP and the Asadpour *et al.* Algorithm

Let $D = (V, A)$ be a digraph. From Section 2.3, recall that a (directed) Hamiltonian circuit in D is a circuit C in D with $V(C) = V$. A function $c\colon A \to \mathbb{R}_+$ is called a *(nonnegative) cost function* on the arcs of D so that $c_a = c(a)$ is called the *cost* of an arc $a \in A$, and $\mathbb{1}_B^\mathsf{T} c$ is called the *cost* of a subset of arcs $B \subseteq A$; moreover, the cost of a subdigraph D' of D is the cost of $A(D')$ (same for a walk in D). The **asymmetric traveling salesman problem** (ATSP) is:

$$\text{Given a digraph } D = (V, A) \text{ and a cost function } c\colon A \to \mathbb{R}_+, \text{ find a Hamiltonian circuit } C \text{ in } D \text{ of minimum cost.} \tag{ATSP}$$

If D is a complete digraph and c is a cost function on the arcs of D, we say c *satisfies* the triangle inequality if for all vertices $u, v, w \in V$ we have $c_{uw} \leq c_{uv} + c_{vw}$. The **metric asymmetric traveling salesman problem** (mATSP) is:

$$\text{Given a complete digraph } D = (V, A) \text{ and a cost function } c\colon A \to \mathbb{R}_+ \text{ that satisfies the triangle inequality, find a Hamiltonian circuit } C \text{ in } D \text{ of minimum cost,} \tag{mATSP}$$

i.e., (mATSP) is the (ATSP) when given a digraph that is complete and a cost function that satisfies the triangle inequality. Note that if (ATSP) and (mATSP) were defined for a cost function taking on reals, the problems would be essentially equal to the ones we present. That is, by just increasing all costs by the absolute value of the lowest cost of an arc, we would have the problems as we present them; they would possibly have distinct optimal values from the former, but the optimum solutions — the Hamiltonian circuits of minimum cost — would still be the same. Hence, we choose to follow a usual presentation where the costs mean, for instance, distances between objects or places.

The **Held-Karp relaxation** of the (ATSP) is the problem: given a digraph $D = (V, A)$ and a cost function $c\colon A \to \mathbb{R}_+$, solve the linear program:

$$\text{Minimize } c^\mathsf{T} x\,, \tag{3.1a}$$

$$\text{subject to } \mathbb{1}_{\delta^{\text{out}}(U)}^\mathsf{T} x \geq 1 \qquad \text{for each } \varnothing \neq U \subsetneq V\,, \tag{3.1b}$$

$$\mathbb{1}_{\delta^{\text{in}}(v)}^\mathsf{T} x = \mathbb{1}_{\delta^{\text{out}}(v)}^\mathsf{T} x = 1 \qquad \text{for each } v \in V\,, \tag{3.1c}$$

$$x \in \mathbb{R}_+^A\,. \tag{3.1d}$$

Let $D = (V, A)$ be a complete digraph, let $c\colon A \to \mathbb{R}_+$ be a cost function, and set $n \coloneqq |V|$. Consider the LP (3.1). Note that $x \coloneqq \frac{1}{n-1}\mathbb{1} \in \mathbb{R}^A$, for instance, is a feasible solution of (3.1) since D is complete; so in this case (3.1) is feasible. Moreover, we can show that (3.1) is bounded. Let x be a feasible solution of (3.1). By (3.1d), we have $x \geq 0$. Also, for each $a \in A$ there exists a unique vertex $v \in V$ such that $a \in \delta^{\text{in}}(v)$, and then, by (3.1c) and as $x \geq 0$, we have $x_a \leq 1$. In other words, $0 \leq x \leq 1$. Thus, by turning (3.1) into the equivalent maximization problem, where we just invert the sign of the cost vector, and by applying Theorem 2.4, we have that the Held-Karp relaxation of (ATSP), when the given digraph is complete, has an optimal solution and a finite optimum value that we denote by $\mathrm{OPT}_{\mathrm{HK}}$.

Now consider the Held-Karp relaxation of the (mATSP) for a complete digraph $D = (V, A)$ and a cost function $c\colon A \to \mathbb{R}_+$ that satisfies the triangle inequality, and let x^* be an optimum solution. Then define the vector z^*, a symmetrized and scaled-down version of x^*, as

$$z^*_{\{u,v\}} \coloneqq \frac{n-1}{n}(x^*_{uv} + x^*_{vu}), \tag{3.2}$$

for each distinct $u, v \in V$. Finally, define the graph $G_{z^*} \coloneqq (V, E)$ where $E \coloneqq \mathrm{supp}(z^*)$, and define the cost function $c^*\colon E \to \mathbb{R}_+$ on the edges of G as

$$c^*_{uv} \coloneqq \min\{c_{uv}, c_{vu}\}, \tag{3.3}$$

for each $\{u, v\} \in E$.

Let U be a finite nonempty set, and let $\{x_i\}_{i\in I}$ be a finite family of points in $\mathbb{R}^U$. A linear combination $\sum_{i\in I} \alpha_i x_i$ is a **convex combination** (of $\{x_i\}_{i\in I}$) if $\alpha_i \geq 0$ for each $i \in I$ and $\sum_{i\in I} \alpha_i = 1$. Then define, for each $X \subseteq \mathbb{R}^U$,

$$\mathrm{conv}\, X \coloneqq \left\{ \sum_{i\in[n]} \lambda_i x_i : n \in \mathbb{N},\ \{x_i\}_{i\in[n]} \subseteq X,\ \lambda \in \mathbb{R}^n_+,\ \text{and} \sum_{i\in[n]} \lambda_i = 1 \right\},$$

called the **convex hull** of X, to be the set of all convex combinations of finite subsets of points of X.

The **spanning tree polytope** of a graph $G = (V, E)$ is

$$P_{\mathrm{sptree}}(G) \coloneqq \mathrm{conv}\{\, \mathbb{1}_F : (V, F) \text{ is a spanning tree of } G\}, \tag{3.4}$$

i.e., the set of all convex combinations of (finite) sets of incidence vectors of spanning trees of G.

Definition 3.1 (α-thin tree)**.** Let $G = (V, E)$ be a graph. Let T be a spanning tree in G, let $\alpha \geq 1$, and let $z \in P_{\mathrm{sptree}}(G)$. Then T is **α-thin** with respect to z if

$$|E(T) \cap \delta(U)| \leq \alpha \mathbb{1}^{\mathsf{T}}_{\delta(U)} z \qquad \text{for each } U \subseteq V. \tag{3.5}$$

Consider the context of Definition 3.1. Note that when $U = \varnothing$ or $U = V$, the condition (3.5) is satisfied for any spanning tree T of G. Thus, to show a spanning tree of G is α-thin with respect to z, we may ignore these cases.

Definition 3.2 ((α, s)-thin tree)**.** Let $D = (V, A)$ be a complete graph, and let $c\colon A \to \mathbb{R}_+$ be a cost function that satisfies the triangle inequality. Let x^* be an optimum solution of the Held-Karp relaxation of (mATSP) determined by D and c, and define z^* from x^* as in (3.2). Moreover, set the graph $G_{z^*} \coloneqq (V, E)$ where $E \coloneqq \mathrm{supp}(z^*)$, and define the cost function $c^*\colon E \to \mathbb{R}_+$ as in (3.3). Finally, set $\mathrm{OPT}_{\mathrm{HK}} \coloneqq c^{\mathsf{T}} x^*$. Then for any $\alpha \geq 1$ and any $s \in \mathbb{R}_+$, a spanning tree T of G_{z^*} is (α, s)-thin if T is α-thin with respect to z^* and

$$\mathbb{1}^{\mathsf{T}}_{E(T)} c^* \leq s \cdot \mathrm{OPT}_{\mathrm{HK}}. \tag{3.6}$$

The following identities will be instrumental in proofs involving x^* and z^*.

Proposition 3.3. Let D, c, x^*, z^*, G_{z^*} be as in Definition 3.2. Then:

(i) Let $u, v \in V$ be distinct, and suppose $\{u, v\} \notin E$. Then $x^*_{uv} = x^*_{vu} = 0$.

(ii) Let U be a subset of V. Then

$$\mathbb{1}^{\mathsf{T}}_{\delta(U)} z^* = \left(1 - \frac{1}{n}\right)\left(\mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}_D(U)} x^* + \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{in}}_D(U)} x^*\right) = 2\left(1 - \frac{1}{n}\right)\mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}_D(U)} x^* = 2\left(1 - \frac{1}{n}\right)\mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{in}}_D(U)} x^*.$$

(iii) Let U be a subset of V. Then $\mathbb{1}^{\mathsf{T}}_{E[U]} z^* = (1 - 1/n)\mathbb{1}^{\mathsf{T}}_{A[U]} x^*$.

Proof. **(i)** Since $E = \operatorname{supp}(z^*)$ and z^* is defined for each two distinct vertices of V, we have $z^*_{\{u,v\}} = 0$. Then $x^*_{uv} + x^*_{vu} = 0$ by (3.2). Moreover, $x^* \geq 0$ by (3.1d). Thus, $x^*_{uv} = x^*_{vu} = 0$.

(ii) By definition of z^* in (3.2) and as $E = \operatorname{supp}(z^*)$, we have

$$\begin{aligned}\mathbb{1}^{\mathsf{T}}_{\delta(U)} z^* &= \left(1 - \frac{1}{n}\right) \sum_{\substack{u \in U, v \notin U \text{ s.t.} \\ x^*_{uv} + x^*_{vu} \neq 0}} (x^*_{uv} + x^*_{vu}) = \left(1 - \frac{1}{n}\right) \sum_{u \in U, v \notin U} (x^*_{uv} + x^*_{vu}) \\ &= \left(1 - \frac{1}{n}\right)\left(\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(U)} x^* + \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(U)} x^*\right) \overset{(3.1c)}{=} 2\left(1 - \frac{1}{n}\right)\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(U)} x^* \overset{(3.1c)}{=} 2\left(1 - \frac{1}{n}\right)\mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(U)} x^* \,.\end{aligned}$$

(iii) By definition of z^* in (3.2), we have

$$\mathbb{1}^{\mathsf{T}}_{E[U]} z^* = \left(1 - \frac{1}{n}\right) \sum_{\{u,v\} \in E[U]} (x^*_{uv} + x^*_{vu}) = \left(1 - \frac{1}{n}\right) \sum_{uv \in A[U]} x^*_{uv} = \left(1 - \frac{1}{n}\right) \mathbb{1}^{\mathsf{T}}_{A[U]} x^* \,,$$

where the second equality holds by item (i). $\square$

The following result shows the graph G_{z^*}, defined as in Definition 3.2, is connected, and so it has a spanning tree. This is a trivial but crucial observation since the second step of the algorithm ApproxATSP by Asadpour, Goemans, Madry, Oveis Gharan, and Saberi [2], as we see next, finds a special spanning tree of G_{z^*}.

Proposition 3.4. Let D, c, x^*, z^*, G_{z^*} be as in Definition 3.2. Then the graph G_{z^*} is connected.

Proof. The proof is by contradiction. Suppose G_{z^*} is disconnected. Then there exists a nonempty and proper subset U of V such that $\delta_{G_{z^*}}(U) = \varnothing$. In other words, for each $u \in U$ and $v \in V \setminus u$ we have $\{u, v\} \notin E$ which, by item (i) of Proposition 3.3, implies $x^*_{uv} = x^*_{vu} = 0$. In particular, this gives $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(U)} x = 0$. However, point x^* is a feasible solution of the Held-Karp relaxation of (mATSP) determined by D and c, and so x^* satisfies (3.1b), a contradiction. $\square$

Now, we present the algorithm due to Asadpour, Goemans, Madry, Oveis Gharan, and Saberi [2] for the metric ATSP.

Algorithm 3.1: ApproxATSP(V, c)

Input:
(i) A finite set of vertices V, where $n \coloneqq |V|$, that define the complete digraph $D \coloneqq (V, A)$.
(ii) A cost function $c\colon A \to \mathbb{R}_+$ that satisfies the triangle inequality.

Output: A Hamiltonian circuit C in D whose cost is $(2\alpha + 2) = O(\ln n / \ln \ln n)$, where $\alpha \coloneqq 4 \ln n / \ln \ln n$, of the optimum value OPT of the (mATSP) determined by D and c.

1. Find an optimal solution x^* for the Held-Karp relaxation of the (mATSP) determined by D and c, define z^* from x^* as in (3.2), and define G_{z^*} as in Definition 3.2
2. Find an (α, s)-thin tree T^* of G_{z^*}, where $\alpha \coloneqq 4 \ln n / \ln \ln n$ and $s \coloneqq 2$, with high probability
3. From T^* find a Hamiltonian circuit C with cost upper bounded by $(2\alpha + s)\mathrm{OPT}_{\mathrm{HK}} \leq (2\alpha + s)\mathrm{OPT}$
4. **return** C

Theorems 3.5, 5.30, and 3.7 show how we can perform the three steps of the algorithm in polynomial time. After that, Theorem 3.8, the main result of the monograph, shows how these three results imply a polynomial-time algorithm that indeed gives a good approximation for the (mATSP) with high probability.

In the next theorem, to analyze the running time of the Held-Karp relaxation of (ATSP), we will need some definitions of Automata theory such as alphabets, words, languages, and size of a word. These are given in Section 6.3 where we also present the optimization and separation problems. Moreover, we will use an encoding scheme, or simply encoding. Roughly speaking, an encoding of a problem is a function that maps the problem instances to strings of an alphabet. In our case, we will map digraphs, represented abstractly with vertex and edge set, to strings of an alphabet with three symbols. The encoding we will choose is a member of a family of encodings that we now introduce. Our presentation is based on Garey and Johnson [10, Sections 1.2, 1.3, and 2.1], where one can find a good description of encodings.

Recall that the time complexity analysis of an algorithm for a problem is done as a function of the input size of the problem. The input size of an instance of a problem, in turn, can only be determined once an encoding is fixed, and different choices of encoding can produce different inputs sizes for the same instance. Therefore, the input size, and so the time complexity analysis of an algorithm will be impacted by the choice of an encoding.

Following principles such as "conciseness" and "decodability", one can design so-called "reasonable" encodings. These principles and what is meant by a "reasoanable" encoding are not formal definitions. However, the encodings that seem to follow those principles also seem to share an interesting property. Although two "reasonable" encodings may produce different input sizes for the same problem instance, they will probably differ at most polynomially from one another. That is, any algorithm that has polynomial-time complexity under one "reasonable" encoding would probably have polynomial-time complexity under all the others; when two encodings have this relation, they are called *polynomially related*. Thus, as long as one chooses a "reasonable" encoding, the polynomial-time complexity of the problem should not be affected.

Therefore, we present and use a recognized "reasonable" encoding for digraphs using adjancecy matrices of digraphs. One can verify this encoding is polynomially related with other standard and "reasoanable" encodings for digraphs, such as listing all vertices and arcs of the digraph, or for each vertex of the digraph listing all the arcs incident to it. In [10, Sections 1.3], Garey and Johnson compare these three encodings for graphs.

The **adjacency matrix** $A_D \in \mathbb{R}^{V \times V}$ of a digraph $D = (V, A)$ without parallel arcs is defined as

$$A_D(i,j) \coloneqq [ij \in A] \qquad \text{for each } i, j \in V\,. \tag{3.7}$$

Set $\mathcal{D}$ to be the set of all (finite) digraphs without parallel arcs. Suppose, without loss of generality, that if $(V, A) \in \mathcal{D}$ with $n \coloneqq |V|$, then $V = \{v_1, v_2, \ldots, v_n\}$. Set $\Sigma \coloneqq \{0, 1, /\}$ to be an alphabet. Define the function $e\colon \mathcal{D} \to \Sigma^*$ as

$$e(D) \coloneqq A_D(v_1,v_1)\ldots A_D(v_1,v_n)/A_D(v_2,v_1)\ldots A_D(v_2,v_n)/\ldots/A_D(v_n,v_1)\ldots A_D(v_n,v_n)\,, \tag{3.8}$$

for each $D \in \mathcal{D}$.

Theorem 3.5. The Held-Karp relaxation of (ATSP) is polynomial-time solvable.

Proof. First, we show that the Held-Karp relaxation of (ATSP) is an optimization problem for a family of polyhedra $(P_\sigma)_{\sigma\in\Pi}$ (see Definition 6.6), for some language Π, that satisfies (6.22). To do that, we show both problems have the same input and the same task.

Set $\mathcal{D}$ to be the set of all (finite) digraphs without parallel arcs. Consider the encoding e of $\mathcal{D}$ using adjacency matrices of digraphs as in (3.8). Set $\Pi \coloneqq e[\mathcal{D}]$. Note that e is injective, and then $e^{-1}[\sigma]$ is an element of $\mathcal{D}$ for each $\sigma \in \Pi$ (see image and preimage definitions in Section 2.2). Thus, we have that any input of the Held-Karp relaxation of (ATSP) can be given by a word $\sigma \in \Pi$ and a cost function $c\colon A \to \mathbb{R}_+$, where $A \coloneqq A(e^{-1}(\sigma))$, so that this problem has the same input of an optimization problem for some family of polyhedra $(P_\sigma)_{\sigma\in\Pi}$.

Now let $\sigma \in \Pi$ be a word, and set $D \coloneqq (V, A) \coloneqq e^{-1}(\sigma)$ to be the corresponding digraph by the encoding e. Also, let $c\colon A \to \mathbb{R}_+$ be a cost function. Note that each string w of Π has the form $w = s_1/s_2/\ldots/s_k$, where k is a positive integer and s_i belongs to $\{0,1\}^{[\ell]}$, for some positive integer ℓ, for each $i \in [k]$; so one can efficiently test whether a string of Σ^* belongs to Π. Also, note that with one traversal of the symbols of σ one can build the set A, and so the set $E_\sigma \coloneqq A$ can be computed from the word σ in time polynomial in $|\sigma|$. Then (6.22a) is satisfied. Moreover, consider the linear program (3.1) for D and c, and then set $P_\sigma \subseteq \mathbb{Q}^A$ to be the polyhedron determined by the system of inequalities (3.1b), (3.1c), and (3.1d). Each inequality of those that determine P_σ has $|A| + 1$ terms with coefficients equal to 0 or 1; thus, each inequality has size upper bounded by a polynomial in $|\sigma|$ as one can see that $|A| \leq |\sigma|$, i.e., P_σ satisfies (6.22b). Therefore, the family $(P_\sigma)_{\sigma\in\Pi}$ satisfies (6.22), and we are done for the first part.

Now we show the optimization problem for the family $(P_\sigma)_{\sigma\in\Pi}$ is polynomial-time solvable. Let us consider the separation problem for the family $(P_\sigma)_{\sigma\in\Pi}$ as given in Definition 6.7. Let $\sigma \in \Pi$, and let $\overline{x} \in \mathbb{Q}^{E_\sigma}$ where $E_\sigma \coloneqq A(e^{-1}[\sigma])$. Also, set $D \coloneqq (V, A) \coloneqq e^{-1}[\sigma]$. To decide whether $\overline{x}$ belongs to P_σ, it is sufficient to test if $\overline{x}$ satisfies the system of inequalities that determine P_σ. The constraint (3.1d) is readily tested in time $O(|A|)$. For constraint (3.1c), first we have $O(|V|)$ steps to range over all vertices of V; then, for each such step, we have to sum some entries of $\overline{x}$. Since for each $a \in A$ there exist exactly two distinct vertices $u, v \in V$ such that $a \in \delta^{\text{in}}(u)$ and $a \in \delta^{\text{out}}(v)$, we have $\sum_{v \in V} |\delta^{\text{in}}(v)| = \sum_{v \in V} |\delta^{\text{out}}(v)| = |A|$, and then there are, in total, $O(|A|)$ entries of $\overline{x}$ being added. Thus, we can test the constraints in (3.1c) in time $O(|V| + |A|)$.

For constraint (3.1b), at a first glance, it seems we would have to test $\Theta(2^{|V|})$ inequalities which, regardless the cost of each inequality, is not done in time $O(p)$ for any polynomial p in $|V|$ and $|A|$. However, if we have computed a vertex set $\varnothing \neq U^* \subsetneq V$ such that $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U^*)}\overline{x} \leq \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)}\overline{x}$ for each $\varnothing \neq U \subsetneq V$, then checking (3.1b) would be reduced to checking just the following inequality

$$\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U^*)}\overline{x} \geq 1\,, \tag{3.9}$$

which can be done in time $O(|A|)$. It turns out we can find a such vertex set in polynomial time. First, note that the edge set $\delta^{\text{out}}(U^*)$ is an s-t cut of minimum capacity, for capacity function $\overline{x}$, in D for any $s \in U^*$ and any $t \in V \setminus U^*$. Second, the problem of, given two distinct vertices s, t of V, finding an s-t cut of minimum capacity, with respect to $\overline{x}$, in D is polynomial in $|V|$ and $|A|$. Thus, to find U^* we just have to find an s-t cut of minimum capacity in D over all distinct vertices s, t in V. More precisely, we can iterate over all distinct vertices s, t in V; then for each such distinct vertices s, t in V we can run the Edmonds-Karp algorithm (see Section 4.2) that returns an s-t cut of minimum capacity in D in time $O(|V||A|^2)$; finally, through all iterations we keep a set U^* such that $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U^*)}\overline{x}$ is minimum. Hence, in time $O(|V|^3|A|^2)$ we find a such set U^* and in time $O(|A|)$ we check (3.9), i.e., in time $O(|V|^3|A|^2)$ we test (3.1b) for $\overline{x}$.

Therefore, testing if $\overline{x} \in P_\sigma$ can be done in time polynomial in $|V|$ and $|A|$ which in turn, as $|V|$ and $|A|$ are polynomials in $|\sigma|$, is a polynomial in $|\sigma|$. In other words, the separation problem for the family $(P_\sigma)_{\sigma \in \Pi}$ is polynomial-time solvable. Thus, by Theorem 6.8, the optimization problem for the family $(P_\sigma)_{\sigma \in \Pi}$ is polynomial-time solvable, and hence the Held-Karp relaxation of (ATSP) is polynomial-time solvable too. $\square$

The following algorithm constitutes the final step in the proof of Theorem 3.7, and ultimately, the final step of Algorithm 3.1. Consider an input D, D', W of Algorithm 3.2. We can loosely denote the arcs of D' as "copies" of the arcs of D, that is, for each arc $a \in A$, the arcs $(a, \beta) \in A \times \mathbb{N}$ are "copies" of a. However, note that there might exist some arc a of D that does not have a "copy" in D'. Also, note that the set A' could be any finite set; the choice of being a subset of $A \times \mathbb{N}$ is made to facilitate the comparison with the arcs of D in Theorem 3.6.

Algorithm 3.2: ShortcutEulerHalmiltonian(D, D', W)

Input:
- (i) a complete digraph $D = (V, A, \psi)$,
- (ii) an Eulerian digraph $D' = (V', A', \psi')$ that is weakly connected and that $V' = V$, $A' \subseteq A \times \mathbb{N}$, and $\psi'((a, \beta)) = \psi(a)$ for each $(a, \beta) \in A'$,
- (iii) a Eulerian cycle $W = (v_0, a_1, v_1, \ldots, a_k, v_k)$ in D'.

Output: A Hamiltonian circuit C in D.

$C \coloneqq (v_0)$
$i \coloneqq 0$ `// During the algorithm,` v_i `will be the last vertex of the current walk` C.
for $j \leftarrow 1$ **to** k **do**
  if $v_j \notin V(C)$ **then**
    Set $a \in A$ such that $\psi(a) = v_i v_j$ `// Such arc exists as` D `is complete and` $v_i \neq v_j$.
    $C \coloneqq C \cdot (v_i, a, v_j)$
    $i \coloneqq j$
Set $a \in A$ such that $\psi(a) = v_i v_k$
return $C \cdot (v_i, a, v_k)$

Theorem 3.6. Let D, D', W be as in the input of Algorithm 3.2. Let $c\colon A \to \mathbb{R}_+$ be a cost function that satisfies the triangle inequality. Set $c'\colon A' \to \mathbb{R}_+$ to be a cost function where $c'((a, \alpha)) \coloneqq c(a)$ for each $(a, \alpha) \in A'$. Then $C \coloneqq \text{ShortcutEulerHalmiltonian}(D, D', W)$ is a Hamiltonian circuit in D with cost upper bounded by the cost of W.

Proof. In algorithm ShortcutEulerHalmiltonian, note that the walk C is built through a sequence of mutations that starts with setting C to (v_0), and then continues with a sequence of concatenations of the current C and a walk of unit length whose vertices lie in W and the arc lies in D.

Since D' is weakly connected and W is an Eulerian trail, we have $V(W) = V$; also, as D is a complete digraph, for any two distinct vertices $u, v \in V$, we have $uv, vu \in \psi(A)$. Thus, we can indeed build a Hamiltonian circuit with the vertices of W and the arcs of D; we show that is the case when we run ShortcutEulerHalmiltonian for D, D', and W.

Let $v \in V$. Set ℓ to be the smallest integer such that $v_\ell = v$. If $\ell = 0$, then v is added to C in Line 1. If $\ell \geq 1$, then in Line 4 we have $v \notin V(C)$, and so v is added to C. Then each vertex of V is added to C at least once.

Suppose there exists $r > \ell$ in $[k]$ such that $v_\ell = v_r = v$. Then in the r iteration of loop in Line 3, the algorithm would consider adding this vertex to C. If $\ell = 0$, then v would have been added to C in Line 1, and if $\ell > 0$, then v would have been added to c in the $\ell < r$ iteration of loop in Line 3; then v would not be added again in C by the condition in Line 4. Thus, all vertices of C, except the last vertex of C, are pairwise distinct. Finally, the first and last vertices of C are v_0. Therefore, C is a Hamiltonian circuit in D.

Now, we show C has cost upper bounded by the cost of W. Set $n \coloneqq |V|$, that is, the length of C as C is a circuit. By construction, $C = (w_0, b_1, w_1 \ldots, b_n, w_n)$ which is defined as: the sequence $(w_0, \ldots, w_{n-1})$ is the subsequence of $(v_0, \ldots, v_k)$ where all occurrences of a vertex in W were removed except its first, and $w_0 = w_n = v_0$; for each $i \in [n]$, we have that b_i is the arc of A such that $\psi(b_i) = w_{j-1}w_j$.

Since the sequence of vertices of C is a subsequence of the sequence of vertices of W and the first and last vertices of C are the first and last vertices of W, respectively, for our goal it suffices to show that the cost of an arbitrary segment (u, b, v) of C is upper bounded by the cost of the corresponding segment of W from u to v (see segment definition in Section 2.3). Then let $i \in [n]$, and set ℓ and r to be the smallest integers such that $v_\ell = w_{i-1}$ and $v_r = w_i$ so that we compare the cost of the segment $P \coloneqq (w_{i-1}, b_i, w_i)$ of C with the cost of the segment $Q \coloneqq (v_\ell, a_{\ell+1}, v_{\ell+1}, \ldots, a_r, v_r)$ of W. Since c satisfies the triangle inequality, we have $c_{b_i} \leq c'_{a_{\ell+1}} + \cdots + c'_{a_r}$, i.e., P has cost upper bounded by the cost of Q. ☐

Theorem 3.7. Let D, c, x^*, z^*, G_{z^*}, c^* be defined as in Definition 3.2. Consider the problems (ATSP) and (3.1), both with respect to D and c, and set OPT and $\mathrm{OPT_{HK}}$ to be their optimal values, respectively. Then there exists a polynomial-time algorithm that, given an (α, s)-thin tree T^* of G_{z^*} for some $\alpha \geq 1$ and $s \in \mathbb{R}_+$, finds a Hamiltonian circuit in D with cost at most $(2\alpha + s)\mathrm{OPT_{HK}} \leq (2\alpha + s)\mathrm{OPT}$.

Proof. Set $\vec{T}^*$ to be an orientation of T^* where each edge of T^* is oriented in the direction of minimum cost, that is, for each edge $\{u, v\} \in E(T^*)$, if $c_{uv} \leq c_{vu}$, then $uv \in A(\vec{T}^*)$, otherwise, $vu \in A(\vec{T}^*)$. Note that, by definition of c^* in (3.3) and definition of $\vec{T}^*$, the cost of T^* equals the cost of $\vec{T}^*$, i.e.,

$$\mathbb{1}^{\mathsf{T}}_{E(T^*)} c^* = \mathbb{1}^{\mathsf{T}}_{A(\vec{T}^*)} c. \tag{3.10}$$

We will augment $\vec{T}^*$ with some arcs with both ends in V so that it becomes an Eulerian digraph D' with cost upper bounded by $(2\alpha + s)\mathrm{OPT_{HK}}$, and then we will build a Hamiltonian circuit C in D, from an Eulerian cycle W in D', with cost upper bounded by the cost of D'. So first, we show we can

$$\begin{gathered}\text{find an Eulerian digraph } D' \text{ on vertex set } V \text{ that contains } \vec{T}^* \text{ as a subdigraph} \\ \text{and has cost upper bounded by } (2\alpha + s)\mathrm{OPT_{HK}}.\end{gathered} \tag{3.11}$$

Recall that a digraph D' is Eulerian if it is weakly connected (i.e., its underlying graph is connected), and if $\mathrm{indeg}_{D'} = \mathrm{outdeg}_{D'}$. Moreover, since the underlying graph of $\vec{T}^*$ is T^*, a tree on vertex set V, any digraph on vertex set V that contains $\vec{T}^*$ as a subdigraph is weakly connected. So, for a digraph D' to be a solution candidate of (3.11), it suffices that

$$D' \text{ has vertex set } V,\ D' \text{ contains } \vec{T}^*, \text{ and } \mathrm{indeg}_{D'} = \mathrm{outdeg}_{D'}. \tag{3.12}$$

Using the condition (3.12), we show how a problem involving nonnegative integral circulations (see circulation definition in (4.24)) solves (3.11).

Let $f \in \mathbb{R}^A_+$ be a nonnegative integer circulation in D. Define the digraph D' *associated* to circulation f as $D' \coloneqq (V', A', \psi')$ where $V' \coloneqq V$, $A' \coloneqq \{\, (a, \beta) \in A \times \mathbb{N} : 1 \leq \beta \leq f_a \}$, and $\psi'((a, \beta)) \coloneqq \psi_D(a)$ for each

$(a, \beta) \in A'$. In other words, the digraph D' has the same vertex set of D, and for each $a \in A$ with $f_a > 0$, it has f_a arcs with the same head and tail as arc a. By [15, (11.3)], each nonnegative integer circulation is the sum of incidence vectors of directed circuits. Hence, the digraph D' can be decomposed in circuits, and then $\mathrm{indeg}_{D'} = \mathrm{outdeg}_{D'}$. Moreover, suppose we identified each arc a of D with the arc $(a, 1)$ of D' so that we can compare D' with $\vec{T}^*$, and define the cost of each arc (a, β) in A' as c_a. Thus, the digraph D' satisfies (3.12) if $f \geq \mathbb{1}_{A(\vec{T}^*)}$, and then if we can find such circulation f with cost upper bounded by $(2\alpha + s)\mathrm{OPT}_{\mathrm{HK}}$, the digraph D' solves (3.11), and we are done for the first part.

A circulation $f \in \mathbb{R}^A_+$, with $f \geq \mathbb{1}_{A(\vec{T}^*)}$, of minimum cost is found by solving the *minimum-cost circulation problem* for the digraph D, the integer lower capacity function $\ell \coloneqq \mathbb{1}_{A(\vec{T}^*)}$, the upper capacity function $u \coloneqq \infty \cdot \mathbb{1}$, and the cost function c, that is,

$$\text{Minimize } c^\mathsf{T} f\,, \tag{3.13a}$$

$$\text{subject to } \mathbb{1}^\mathsf{T}_{\delta^{\mathrm{in}}_D(v)} f = \mathbb{1}^\mathsf{T}_{\delta^{\mathrm{out}}_D(v)} f \qquad \text{for each } v \in V\,, \tag{3.13b}$$

$$f_a \geq \ell_a \qquad \text{for each } a \in A\,, \tag{3.13c}$$

By [15, Corollary 12.2a.], there is a polynomial-time algorithm that solves this problem and, since ℓ, u are integral, finds an integer circulation as an optimum solution. Hence, it only remains to show its optimal value is upper bounded as desired. If we look at the following subset of the feasible region of (3.13)

$$\{\, f \in \mathbb{R}^A : f \text{ is a circulation, and } \ell \leq f \leq u \coloneqq \ell + 2\alpha x^* \}\,, \tag{3.14}$$

we have that any circulation f in the set from (3.14) will have the desired cost, and so an optimal solution f^* of (3.13) too, that is,

$$\begin{aligned} c^\mathsf{T} f^* &\leq c^\mathsf{T} f \leq c^\mathsf{T} u = c^\mathsf{T}(\ell + 2\alpha x^*) = c^\mathsf{T}\mathbb{1}_{A(\vec{T}^*)} + 2\alpha c^\mathsf{T} x^* \overset{(3.10)}{=} \mathbb{1}^\mathsf{T}_{E(T^*)} c^* + 2\alpha c^\mathsf{T} x^* \overset{(3.6)}{\leq} s \cdot \mathrm{OPT}_{\mathrm{HK}} + 2\alpha c^\mathsf{T} x^* \\ &= (2\alpha + s)\mathrm{OPT}_{\mathrm{HK}}\,. \end{aligned}$$

So we reduce our task to show the set in (3.14) is nonempty. By Hoffman's Circulation Theorem (see Theorem 4.16), it is necessary and sufficient to show that

$$\mathbb{1}^\mathsf{T}_{\delta^{\mathrm{in}}_D(S)} \ell \leq \mathbb{1}^\mathsf{T}_{\delta^{\mathrm{out}}_D(S)} u \qquad \text{for each } S \subseteq V\,. \tag{3.15}$$

Thus, first note that $\mathbb{1}^\mathsf{T}_{\delta^{\mathrm{in}}_D(S)} \ell = |A(\vec{T}^*) \cap \delta^{\mathrm{in}}_D(S)|$, and set $G \coloneqq G_{z^*}$. Second, if $vw \in A(\vec{T}^*)$, then $\{v, w\}$ is an edge of T^* and G, since $\vec{T}^*$ is an orientation of T^* and T^* is a subgraph of G; so $vw \in A(\vec{T}^*) \cap \delta^{\mathrm{in}}_D(S)$ implies $\{v, w\} \in E(T^*) \cap \delta_G(S)$. Also, if $vw \in \delta^{\mathrm{in}}_D(S)$, then $wv \notin \delta^{\mathrm{in}}_D(S)$. Hence,

$$\begin{aligned} |A(\vec{T}^*) \cap \delta^{\mathrm{in}}_D(S)| &\leq |E(T^*) \cap \delta_G(S)| \\ &\leq \alpha \mathbb{1}^\mathsf{T}_{\delta_G(S)} z^* && \text{as } T^* \text{ is an } \alpha\text{-thin tree (see Definition 3.1)} \\ &= \alpha\left(1 - \frac{1}{n}\right)\left(\mathbb{1}^\mathsf{T}_{\delta^{\mathrm{in}}_D(S)} x^* + \mathbb{1}^\mathsf{T}_{\delta^{\mathrm{out}}_D(S)} x^*\right) && \text{by item (ii) of Proposition 3.3} \\ &= \alpha\left(1 - \frac{1}{n}\right)\left(2\mathbb{1}^\mathsf{T}_{\delta^{\mathrm{out}}_D(S)} x^*\right) && \text{by (3.1c)} \\ &< \mathbb{1}^\mathsf{T}_{\delta^{\mathrm{out}}_D(S)} u && \text{as } u = \ell + 2\alpha x^* \text{ with } \ell \geq 0 \text{ and } (1 - 1/n) < 1\,. \end{aligned}$$

Therefore, (3.14) is nonempty, and we can solve (3.11), and in polynomial-time.

Now let f^* be an integer optimum solution of (3.13), and set D' to be the digraph associated to f^*. Let W be a Eulerian cycle in D' that can be found in polynomial time by, for instance, Fleury's algorithm. We *shortcut* W into a walk $C \coloneqq \mathrm{ShortcutEulerHalmiltonian}(D, D', W)$ (see Algorithm 3.2), and then, by Theorem 3.6, C is a Hamiltonian circuit in D of cost upper bounded by the cost of W, which is the cost of D' since W is an Eulerian trail in D'. □

Theorem 3.8 (Main Result)**.** Let $D = (V, A)$ be a complete digraph, and let $c\colon A \to \mathbb{R}_+$ be a cost function that satisfies the triangle inequality. Set $n \coloneqq |V|$. Then $\mathrm{ApproxATSP}(V, c)$ is a $(2+8\ln n/\ln\ln n)$-approximate solution to the instance of (mATSP) determined by D and c, with high probability, and in time polynomial in the size of the input.

Proof. By Theorem 3.5, we find in Line 1, in polynomial time, an optimal solution x^* for the Held-Karp relaxation of the (mATSP) determined by D and c. Then define z^* from x^* as in (3.2), and define G_{z^*} as in Definition 3.2.

By Theorem 5.30, we find in Line 2, in polynomial time and with high probability, an $(4\ln n/\ln\ln n, 2)$-thin tree T^* of G_{z^*}.

Finally, by applying Theorem 3.7 to T^*, we find in Line 3, in polynomial time, a Hamiltonian circuit C in D that is a $(2 + 8\ln n/\ln\ln n)$-approximate solution to the instance of (mATSP) defined by D and c.

Moreover, since each of the three lines of Algorithm 3.1 is done in polynomial time, the whole algorithm takes polynomial time. $\square$

Chapter 4

Max-Flow Min-Cut and Circulations

4.1 The Max-Flow Min-Cut Theorem

Let $D = (V, A, \psi)$ be a digraph, and let s and t be distinct vertices of V. A function $f \colon A \to \mathbb{R}$ is an s-t **flow** if it satisfies the following conditions:

$$f(a) \geq 0 \qquad \text{for each } a \in A\,, \text{ and} \tag{4.1a}$$

$$\mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}(v)} f = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(v)} f \qquad \text{for each } v \in V \setminus \{s, t\}\,. \tag{4.1b}$$

Let us denote the net amount of flow *entering* a subset of vertices by the **excess** function $\operatorname{excess}_f \colon \mathcal{P}(V) \to \mathbb{R}$ defined by $\operatorname{excess}_f(U) \coloneqq \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}(U)} f - \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)} f$ for each $U \subseteq V$. For simplicity, set $\operatorname{excess}_f(v) \coloneqq \operatorname{excess}_f(\{v\})$ for any $v \in V$.

The net amount of flow *leaving* s is called the **value** of an s-t flow f, and it is defined as $\operatorname{value}(f) \coloneqq -\operatorname{excess}_f(s)$. The following proposition presents two identities regarding the arcs incident to a subset of vertices of V. We will use this proposition to prove Theorem 4.2.

Proposition 4.1. Let $D = (V, A, \psi)$ be a digraph, and let $U \subseteq V$. Then

$$\sum_{u \in U} \mathbb{1}_{\delta^{\text{in}}_D(u)} = \mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{in}}_D(U)}\,, \text{ and} \tag{4.2}$$

$$\sum_{u \in U} \mathbb{1}_{\delta^{\text{out}}_D(u)} = \mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{out}}_D(U)}\,. \tag{4.3}$$

Proof. Let $a \in A$. Since $\delta^{\text{in}}_D(u) \cap \delta^{\text{in}}_D(v) = \varnothing$ for each $u, v \in V$ distinct, it follows that

$$\begin{aligned}
\left(\sum_{u \in U} \mathbb{1}_{\delta^{\text{in}}_D(u)}\right)_a &= [a \in \delta^{\text{in}}_D(w) \text{ for some } w \in U] \\
&= [(\text{there exist } v \in U \text{ and } w \in U \text{ such that } \psi(a) = (v, w)) \text{ or} \\
&\qquad (\text{there exist } v \in \overline{U} \text{ and } w \in U \text{ such that } \psi(a) = (v, w))] \\
&= [(a \in A[U]) \text{ or } (a \in \delta^{\text{in}}_D(U))] \\
&= \left(\mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{in}}_D(U)}\right)_a\,,
\end{aligned}$$

where the last equality holds since $A[U] \cap \delta^{\text{in}}_D(U) = \varnothing$. This proves (4.2).

Now let $D^{-1} = (V, A, \psi^{-1})$ be the reverse digraph of D, where $\psi^{-1}(a) = \psi(a)^{-1}$ for each $a \in A$ (recall definition in Section 2.3). Then $\delta^{\text{out}}_D(u) = \delta^{\text{in}}_{D^{-1}}(u)$ for each $u \in U$, and so $\delta^{\text{out}}_D(U) = \delta^{\text{in}}_{D^{-1}}(U)$. Moreover, note that for each $a \in A$, $\psi(a) \in U \times U$ iff $\psi^{-1}(a) \in U \times U$, and hence $A[U] = A^{-1}[U]$. Thus,

$$\sum_{u \in U} \mathbb{1}_{\delta^{\text{out}}_D(u)} = \sum_{u \in U} \mathbb{1}_{\delta^{\text{in}}_{D^{-1}}(u)} = \mathbb{1}_{A^{-1}[U]} + \mathbb{1}_{\delta^{\text{in}}_{D^{-1}}(U)} = \mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{out}}_D(U)}\,,$$

where the second equality holds by (4.2) applied to D^{-1}. This proves (4.3). $\square$

The next result regards the excess function excess_f for any function $f : A \to \mathbb{R}$, where A is the set of arcs of a digraph. When we applied the result to an s-t flow of a digraph $D = (V, A, \psi)$ we discover that: the net amount of flow entering a subset U of V equals the sum, for each vertex u of U, of the net amount of flow entering u.

Theorem 4.2. Let $D = (V, A, \psi)$ be a digraph, and let $f \colon A \to \mathbb{R}$. Then for every $U \subseteq V$ we have

$$\operatorname{excess}_f(U) = \sum_{u \in U} \operatorname{excess}_f(u)\,. \tag{4.4}$$

Proof. By definition of excess function, we have

$$\begin{aligned}\sum_{u\in U}\operatorname{excess}_f(u) &= \sum_{u\in U}\big(\mathbb{1}_{\delta^{\text{in}}(u)} - \mathbb{1}_{\delta^{\text{out}}(u)}\big)^{\mathsf{T}} f = \Bigg(\sum_{u\in U}\big(\mathbb{1}_{\delta^{\text{in}}(u)} - \mathbb{1}_{\delta^{\text{out}}(u)}\big)\Bigg)^{\mathsf{T}} f\\ &= \Bigg(\sum_{u\in U}\mathbb{1}_{\delta^{\text{in}}(u)} - \sum_{u\in U}\mathbb{1}_{\delta^{\text{out}}(u)}\Bigg)^{\mathsf{T}} f = \Big(\big(\mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{in}}(U)}\big) - \big(\mathbb{1}_{A[U]} + \mathbb{1}_{\delta^{\text{out}}(U)}\big)\Big)^{\mathsf{T}} f\\ &= \big(\mathbb{1}_{\delta^{\text{in}}(U)} - \mathbb{1}_{\delta^{\text{out}}(U)}\big)^{\mathsf{T}} f = \operatorname{excess}_f(U),\end{aligned}$$

where the fourth equality follows from Proposition 4.1. $\square$

The equation (4.4) of last theorem reveals two interesting properties of the excess function. To explain that, let $D = (V, A, \psi)$ be a digraph, let $f \colon A \to \mathbb{R}$ be a function (f is not necessarily an s-t flow for $s, t \in V$ distinct), and let excess_f be an excess function. On the one hand, excess_f can have a compact representation. For instance, the image of excess_f for just $|V|$ points suffices to compute its image for all $2^{|V|}$ points of its domain. On the other hand, for each $U \subseteq V$, to compute $\sum_{u\in U}\operatorname{excess}_f(u)$ is reduced to compute $\operatorname{excess}_f(U)$. In other words, to compute $\sum_{u\in U}\operatorname{excess}_f(u)$ for each $U \subseteq V$, instead of taking into account the image of f for each arc that enters or leaves a vertex $u \in U$, which would comprise the arcs with both ends in U, that is, $A[U]$, and the arcs with exactly one end in U, that is, $\delta^{\text{in}}(U) \cup \delta^{\text{out}}(U)$, we can just look at the image of f for the arcs with exactly one end in U.

Besides these properties of the excess function, when applied to s-t flows of a digraph, Theorem 4.2 provides interesting consequences that we will see in the next four corollaries, which, in turn, underlie the main results of the section and chapter. First, for a digraph $D = (V, A, \psi)$, vertices $s, t \in V$ distinct, an s-t flow f, and an excess function excess_f, if we set $U := V$, then we shall conclude that there exists a vertex $u \in V$ with $\operatorname{excess}_f(u) > 0$ iff there exists a vertex $v \in V$ with $\operatorname{excess}_f(v) < 0$. Equivalently, we shall conclude that either the excess function is identically zero or it has entries with both signs (and they cancel out since $\operatorname{excess}_f(V) = 0$). This result is going to be useful soon, for example, in the proof of Theorem 4.16, the Hoffman's circulation theorem.

Corollary 4.3. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $f \colon A \to \mathbb{R}$ be an s-t flow. Set $S := \{\, v \in V : \operatorname{excess}_f(v) > 0\}$ and $T := \{\, v \in V : \operatorname{excess}_f(v) < 0\}$. Then either $\operatorname{excess}_f = 0$ or $S \neq \varnothing \neq T$.

Proof. Suppose $S \neq \varnothing$ and $T = \varnothing$. Then $\sum_{v\in V}\operatorname{excess}_f(v) = \sum_{v\in S}\operatorname{excess}_f(v) + \sum_{v\notin S}\operatorname{excess}_f(v) > 0$. However, by Theorem 4.2, we have $\sum_{v\in V}\operatorname{excess}_f(v) = \operatorname{excess}_f(V) = 0$. The proof for $S = \varnothing$ and $T \neq \varnothing$ is similar. $\square$

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let f be an s-t flow in D. The following two corollaries provide two alternative forms to compute $\operatorname{value}(f)$. These forms will come in handy when we analyze our main problem of trying to increase $\operatorname{value}(f)$ with f subject to some constraints. Since $\operatorname{value}(f) = -\operatorname{excess}_f(s)$, the next corollary, in particular, says the value of f equals the net amount of flow entering t.

Corollary 4.4. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $f \colon A \to \mathbb{R}$ be an s-t flow. Then

$$\operatorname{excess}_f(s) = -\operatorname{excess}_f(t)\,.$$

Proof. Set $U := V \setminus \{t\}$. On the one hand, $\operatorname{excess}_f(s) = \sum_{u\in U} \operatorname{excess}_f(u)$ by the flow conservation condition (4.1b). On the other hand, by Theorem 4.2, we have $\operatorname{excess}_f(U) + \operatorname{excess}_f(t) = \operatorname{excess}_f(V) = 0$, so $-\operatorname{excess}_f(t) = \operatorname{excess}_f(U)$. Thus, the result comes from applying the equation of Theorem 4.2 for this set U. ☐

Let $D = (V, A, \psi)$ be a digraph, and let $s, t \in V$ be distinct. A subset B of A is called an s-t **cut** if $B = \delta^{\text{out}}(U)$ for some $U \subset V$ with $s \in U$ and $t \notin U$. The following result shows that for any s-t flow f in D and for any $U \subset V$ that defines an s-t cut, the value of f equals the net amount of flow leaving U.

Corollary 4.5. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, let $f\colon A \to \mathbb{R}$ be an s-t flow, and let $U \subset V$ with $s \in U$ and $t \notin U$. Then $\operatorname{value}(f) = -\operatorname{excess}_f(U)$.

Proof. We have $\operatorname{value}(f) = -\operatorname{excess}_f(s)$ by the definition of value of f. On the other hand, we have $-\operatorname{excess}_f(s) = -\operatorname{excess}_f(U)$ by the flow conservation condition (4.1b) and by Theorem 4.2. ☐

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. An s-t flow f is said to be **under** u or **subject to** u if $f \le u$, and u is usually called a capacity function. For $B \subseteq A$, it is usual to call $\mathbb{1}_B^{\mathsf{T}} u$ the capacity of B. Thus, we see next that the value of any s-t flow is bounded above by the capacity of any s-t cut.

Corollary 4.6. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Then for any s-t flow f subject to u and any $U \subset V$ with $s \in U$ and $t \notin U$ it holds that

$$\operatorname{value}(f) \le \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)} u\,.$$

Proof. By Corollary 4.5, we have

$$\operatorname{value}(f) = -\operatorname{excess}_f(U) = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)} f - \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}(U)} f \le \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)} f \le \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(U)} u\,,$$

where the first inequality follows from non-negativity (4.1a) of f and $\mathbb{1}_{\delta^{\text{in}}(U)} \ge 0$, and the second one from the fact that $f \le u$ and $\mathbb{1}_{\delta^{\text{out}}(U)} \ge 0$. ☐

By this point, we have introduced the excess function and some of its properties and, from this function, we find out some relations involving the value of an s-t flow. The latter relation represents the first step in establishing the best upper bound possible for the value of an s-t flow that will culminate in the max-flow min-cut theorem. This concerns the problem we now introduce. A **maximum flow** is an s-t flow f subject to u of maximum value, and the problem of finding such flow in a digraph is called the **maximum flow problem**. Its formulation as an optimization problem, i.e., with an objective function and a set of constraints that define its feasible region, is as follows:

$$\begin{aligned}
&\text{Maximize } \operatorname{value}(f)\,, && && (4.5a)\\
&\text{subject to } \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}(v)} f = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}(v)} f && \text{for each } v \in V \setminus \{s,t\}\,, && (4.5b)\\
&\qquad f \le u\,, && && (4.5c)\\
&\qquad f \in \mathbb{R}^A_+\,, && && (4.5d)
\end{aligned}$$

and it provides a pleasant surprise as we now see. First, notice that the feasible region is determined only by non-strict linear inequalities and equations on f. Besides, the objective function is also linear in f. Thus, (4.5) is actually a *linear program* as we define in (2.10).

The following result shows that there always exist an s-t flow of maximum value. We could prove that using a result from real analysis, due to Weierstrass, that says: if $g\colon X \to \mathbb{R}$ is a continuous function, and if X is a nonempty, closed and bounded subset of $\mathbb{R}^n$, then there exist $x_1, x_2 \in X$ such that $g(x_1) \le g(x) \le g(x_2)$ for each $x \in X$. In the case of the maximum flow problem (4.5) with digraph $D = (V, A, \psi)$, the function would be $f \in \{\, f \in \mathbb{R}^A : f \text{ is an } s\text{-}t \text{ flow in } D \text{ under } u\} \mapsto \operatorname{value}(f) \in \mathbb{R}$. However, we prove the existence of a maximum flow, taking advantage of the linear programming formulation of the problem.

Proposition 4.7. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Then there exists an s-t flow f in D subject to u of maximum value.

Proof. We look at (4.5). By Corollary 4.6, the problem (4.5) is bounded. Besides, its feasible region is nonempty, since $f = 0$ is a feasible solution. Therefore, as a result of being a bounded and feasible linear program, by Theorem 2.4, the maximum flow problem has an optimum solution. $\square$

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u \colon A \to \mathbb{R}_+$. Let $f \colon A \to \mathbb{R}$ be an s-t flow subject to u, and set $n := |V|$. Now that we know a maximum flow always exists, the problem is how to find one, and efficiently. We develop here an idea involving what we call pushing flow through s-t paths in D in order to increase the value of f. Recall, from Corollary 4.4, that value(f) equals the net amount of flow entering t. So in essence, we will try to increase the amount of flow entering t by searching for s-t paths in D in which we can increase f while f continues to be an s-t flow subject to u. More precisely, let $P = (s, a_1, \dots, a_k, t)$ be an s-t path in D for $k \in [n-1]$, and set $\varepsilon := \min_{i \in [k]}\{u(a_i) - f(a_i)\}$. If $\varepsilon > 0$, then we can increase f, in each arc of $A(P)$, by the same positive value, smaller than or equal to ε, so that f is still an s-t flow in D and the amount of flow entering t increases. This seems a good start; however, if we look at the two digraphs in Figure 4.1, we cannot increase the current s-t flows by this approach but still they are not maximum.

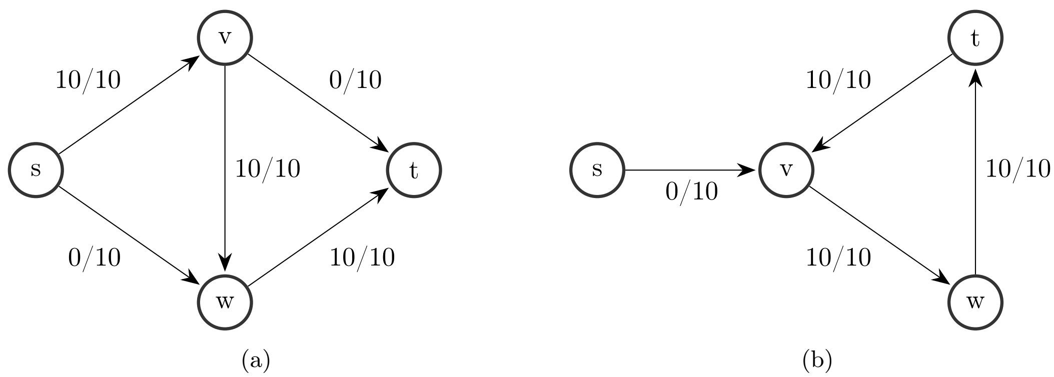

Figure 4.1: Unable to push flow through s-t paths, but still both current s-t flows are not maximum

In the digraph of Figure 4.1a, we could increase 10 units of flow in (s, w) while redirecting the flow from (v, w) to (v, t). In the digraph of Figure 4.1b, we could decrease 10 units of flow leaving from t while, in order to maintain the flow conservation condition (4.1b) in v, increasing the flow in (s, v) by 10 units. In both cases, we would increase the value of the flow by 10 units, and since this would saturate the capacity of the arcs leaving s, by Corollary 4.6, we would have a maximum flow. This suggests we should consider decreasing flow from arcs in our strategy to find a maximum flow. So is our first approach totally doomed? Not exactly. We introduce next a new digraph where pushing flow through s-t paths will be associated with pushing and/or decreasing flow of the corresponding s-t paths in the original digraph, and hence applying our first approach in the new digraph will lead us to a maximum flow in the original digraph.

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, let $u \colon A \to \mathbb{R}_+$, and let $f \colon A \to \mathbb{R}$ be an s-t flow subject to u. Let $D_f = (V, A_f, \psi_f)$ be the digraph such that $A_f := F \cup B$ where

$$F := \{\, (a, +1) \in A \times \{+1\} : f(a) < u(a)\} \text{ and } B := \{\, (a, -1) \in A \times \{-1\} : f(a) > 0\}\,, \tag{4.6}$$

and the incidence function is defined by

$$\psi_f((a, \alpha)) := \psi(a)^{\alpha}\,, \tag{4.7}$$

for each $(a, \alpha) \in A_f$ with $a \in A$ and $\alpha \in \{-1, +1\}$ (recall that $(v, w)^{-1} = (w, v)$ for any $v, w \in V$). This digraph is called the **residual digraph** of D with respect to f and u (the digraph D and the capacity function u usually can be deduced from the definition of the s-t flow f; in that case, we just say residual digraph of f). The capacity function $u_f \colon A_f \to \mathbb{R}_+$ associated with D_f is called **residual capacity** and defined by

$$u_f((a, \alpha)) := \begin{cases} u(a) - f(a)\,, & \text{if } \alpha = +1\,, \\ f(a)\,, & \text{if } \alpha = -1\,, \end{cases} \tag{4.8}$$

for each $(a, \alpha) \in A_f$ with $a \in A$ and $\alpha \in \{-1, +1\}$. Note that D_f has at most twice many arcs as D since the set of arcs A_f of D_f is a subset of $A \times \{-1, +1\}$.

Now we detail the relation between the arcs of A and A_f, and we establish how pushing flow in an arc of D_f changes f. Let $a \in A$. If $f(a) < u(a)$, then we can push at most $u(a) - f(a)$ units of flow in a so that f remains subject to u. We represent that possibility by the arc $(a, +1)$ in F, the set of so-called *forward* arcs, with $\psi_f((a, +1)) = \psi(a)$, and with capacity $u_f((a, +1)) = u(a) - f(a)$. Then we establish that when we push at most $u_f(a)$ units of flow in $(a, +1)$ to that corresponds increasing f in a by the same amount. At the same time, if $f(a) > 0$, then we can decrease at most $f(a)$ units of flow in a so that f remains non-negative. We represent that possibility by the arc $(a, -1)$ in B, the set of so-called *backward* arcs, with $\psi_f((a, -1)) = \psi(a)^{-1}$, and with capacity $u_f((a, -1)) = f(a)$. Then we establish that when we push at most $u_f(a)$ units of flow in $(a, -1)$ to that corresponds decreasing f in a by the same amount.

Therefore, with the residual digraph D_f we can not only represent how much f can change, increasing or decreasing in each arc of D, so that f remains non-negative and subject to u but also, by *just pushing* flow in the arcs of D_f, increase or decrease f in each arc of D. However, when we modify f, we also want its flow conservation condition (4.1b) satisfied so that f indeed remains an s-t flow in D, and we want that the value of f increases. We claim that is what happens with f when we push flow through s-t paths in D_f. We illustrate that with the two digraphs of Figure 4.1.

In Figure 4.1a, denote by $D = (V, A)$ the digraph, by f the current s-t flow in D, and by u the capacity function. From Section 2.3, recall that as we state the digraph D as a pair, we have that A is a subset of $V \times V$ and the omitted incidence function is the identity function. Consider the residual digraph D_f of f. By definition of residual digraph, since the flow in arcs (s, w) and (v, t) is smaller than their capacities, and the flow in arc (v, w) is bigger than zero, we have that $a_1 := ((s, w), +1)$ and $a_3 := ((v, t), +1)$ are forward arcs in D_f while $a_2 := ((v, w), -1)$ is a backward arc in D_f; arcs a_1, a_2, a_3 have 10 units of residual capacity. Thus, $P := (s, a_1, w, a_2, v, a_3, t)$ is an s-t path in D_f. Suppose we push 10 units of flow in P. Then f increases 10 units in arcs (s, w) and (v, t) and decreases 10 units in arc (v, w). By definition of residual capacity, f continues to be non-negative and subject to u. Moreover, in vertex v the amount of flow that increases entering by (s, w) is the same that decreases entering by (v, w), and in vertex w the amount of flow that decreases leaving by (v, w) is the same that increases leaving by (v, t), i.e., the flow conservation (4.1b) for f is satisfied. Finally, the amount of flow entering t increases by 10 units. In other words, f continues to be an s-t flow in D and its value has increased.

In Figure 4.1b, we make a similar analysis. Again, denote the digraph by $D = (V, A)$, the current s-t flow in D by f, the capacity function by u, and consider the residual digraph D_f of f. Since $f((s, v)) < u((s, v))$ and $f((t, v)) > 0$, we have that $a_1 := ((s, v), +1)$ is a forward arc in D_f while $a_2 := ((t, v), -1)$ is a backward arc in D_f; both a_1, a_2 have 10 units of residual capacity. Thus, $P := (s, a_1, v, a_2, t)$ is an s-t path in D_f. Suppose we push 10 units of flow in P. Then f increases 10 units in arcs (s, v) and decreases 10 units in arc (t, v). Again by definition of residual capacity, f continues to be non-negative and subject to u; in v the amount of flow that increases entering by (s, v) is the same amount that decreases entering by (t, v) while in w the amount of flow entering or leaving has not changed, i.e., the flow conservation condition (4.1b) for f is satisfied; the amount of flow leaving t decreases by 10 units. Therefore, as before, f continues to be an s-t flow in D and its value has increased.

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u \colon A \to \mathbb{R}_+$. Let f be an s-t flow in D subject to u, and let $D_f = (V, A_f, \psi_f)$ be the residual digraph of f. Those cases we analyze in Figure 4.1 are coherent with our claim that pushing flow through s-t paths in D_f increase the value of f. However, is that true for any s-t path in D_f? For instance, if P is an s-t path in D_f with at least three vertices, then each vertex v in P that is neither s nor t can have either a forward or a backward arc in P entering v as well as either a forward or a backward arc in P leaving v; this amounts to four cases to be analyzed regarding flow conservation (4.1b). In the next Proposition 4.8, we prove our claim for any s-t path in D_f. Before, we must make a remark. Although we have mentioned pushing flow in the residual digraph, we will not actually work with an s-t flow in the residual digraph. That was only a mean for illustration. In fact, we are going to use the residual digraph only to decide how to modify an s-t flow in D. To do that we introduce the following notation, and Proposition 4.8 should clarify its use.

For each s-t path P in D_f, we define $\operatorname{dir}_P \in \mathbb{R}^A$ by

$$\operatorname{dir}_P(a) := \begin{cases} +1\,, & \text{if } (a,+1) \in A(P)\,,\\ -1\,, & \text{if } (a,-1) \in A(P)\,,\\ 0\,, & \text{otherwise}\,, \end{cases} \tag{4.9}$$

for each $a \in A$. Note that even if $(a,-1)$ and $(a,+1)$ are in A_f, since P is a path in D_f, they both cannot be together in P. Thus, dir_P is indeed a function.

Proposition 4.8. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Let f be an s-t flow in D subject to u, and let $D_f = (V, A_f, \psi_f)$ be the residual digraph of f with residual capacity function u_f. If P is an s-t path in D_f and $0 < \varepsilon \leq u_f(a)$ for each $a \in A(P)$, then $f' := f + \varepsilon \cdot \operatorname{dir}_P$ is an s-t flow in D subject to u. Moreover, $\operatorname{value}(f') = \operatorname{value}(f) + \varepsilon > \operatorname{value}(f)$.

Proof. First we show both non-negativity (4.1a) of f' and $f' \leq u$. Let $a \in A$. If both $(a,-1)$ and $(a,+1)$ are not in $A(P)$, then $\operatorname{dir}_P(a) = 0$, and hence $f'(a) = f(a)$ and $0 \leq f'(a) \leq u(a)$. If $(a,-1)$ is a backward arc in $A(P)$, then $\operatorname{dir}_P(a) = -1$ and

$$\begin{aligned} f'(a) &= f(a) - \varepsilon \geq f(a) - u_f((a,-1)) = f(a) - f(a) = 0\,, \text{ and}\\ f'(a) &= f(a) - \varepsilon \leq f(a) \leq u(a)\,. \end{aligned}$$

If $(a,+1)$ is a forward arc in $A(P)$, then $\operatorname{dir}_P(a) = +1$ and

$$\begin{aligned} f'(a) &= f(a) + \varepsilon > f(a) \geq 0\,, \text{ and}\\ f'(a) &= f(a) + \varepsilon \leq f(a) + u_f((a,+1)) = f(a) + (u(a) - f(a)) = u(a)\,. \end{aligned}$$

Now for f' we show flow conservation condition (4.1b) for each vertex of V except for s and t. Let $v \in V \setminus \{s,t\}$. We show $\operatorname{excess}_{f'}(v) = 0$. Suppose $v \notin V(P)$, and let a be an arc of D incident to v. Since $v \notin V(P)$, it follows $(a,-1)$ and $(a,+1)$ are not in $A(P)$, and hence $\operatorname{dir}_P(a) = 0$ with $f'(a) = f(a)$. Thus, $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(v)} f' = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(v)} f$ and $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(v)} f' = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(v)} f$, and then $\operatorname{excess}_{f'}(v) = \operatorname{excess}_f(v) = 0$.

Now suppose $v \in V(P)$. Then there exists an arc (b,β) of P entering v and an arc (c,γ) of P leaving v. We prove that they together do not change $\operatorname{excess}_f(v)$, i.e., that $\operatorname{excess}_{f'}(v) = 0$. Let an arc $(a,\alpha) \in A(P)$. If $\alpha = +1$ (forward case), then $f'(a) = f(a) + \varepsilon$. On the other hand, if $\alpha = -1$ (backward case), then $f'(a) = f(a) - \varepsilon$. Besides, we have $\operatorname{excess}_f(v) = 0$ by flow conservation condition (4.1b) for f. Thus,

$$\begin{aligned} \operatorname{excess}_{f'}(v) &= \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(v)} f' - \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(v)} f'\\ &= \Big(\mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(v)} f + \Big[(b,+1) \in \delta^{\text{in}}_{A(P)}(v)\Big]\varepsilon - \Big[(c,-1) \in \delta^{\text{out}}_{A(P)}(v)\Big]\varepsilon\Big)\\ &\quad - \Big(\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(v)} f - \Big[(b,-1) \in \delta^{\text{in}}_{A(P)}(v)\Big]\varepsilon + \Big[(c,+1) \in \delta^{\text{out}}_{A(P)}(v)\Big]\varepsilon\Big). \end{aligned}$$

For each of the four possibilities of arc entering and arc leaving v in P (both (b,β) and (c,γ) can be a forward or a backward arc), it holds from the last equation that $\operatorname{excess}_{f'}(v) = \operatorname{excess}_f(v) + \varepsilon - \varepsilon = 0$.

Finally, we show $\operatorname{value}(f') = \operatorname{value}(f) + \varepsilon$. First, by definition, we have that $\operatorname{value}(f) = -\operatorname{excess}_f(s) = \mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(s)} f - \mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(s)} f$. Since P is a path and s is the first vertex of P, then s has only one arc (a,α), with $\alpha \in \{-1,+1\}$, incident to it in P and it is one leaving s. If $\alpha = -1$, then $a \in \delta^{\text{in}}_D(s)$ and it decreases $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{in}}_D(s)} f$ by ε. If $\alpha = +1$, then $a \in \delta^{\text{out}}_D(s)$ and it increases $\mathbb{1}^{\mathsf{T}}_{\delta^{\text{out}}_D(s)} f$ by ε. Either way, we would increase $\operatorname{value}(f')$ by ε with respect to $\operatorname{value}(f)$ as desired. ☐

Although our ultimate goal in this section is to prove Theorem 4.10, the Max-flow Min-Cut theorem, we can consider the last 3 results of this section by the perspective of creating an algorithm for the maximum flow problem. In the next section, the algorithm we shall build is based on two facts: one can increase the value of a flow by pushing flow in the corresponding residual digraph, which is proved by Proposition 4.8; a criterion by which the algorithm can assert a flow is maximum and then terminate, that is, a flow is maximum if and only if the corresponding residual digraph has no s-t paths. Proposition 4.9 shows that "no s-t paths in the residual digraph" is a sufficient condition for the corresponding flow to be maximum while Theorem 4.10 shows it to be a necessary condition.

Moreover, we should notice that an interesting s-t cut is discovered in a digraph when the corresponding residual digraph has no s-t paths, one that has minimum capacity among all s-t cuts of the digraph, which is called minimum cut. Thus, we shall tie together the ideas of maximum flow, minimum cut, and residual digraph with no s-t paths.

Proposition 4.9. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, let $u\colon A \to \mathbb{R}_+$. Let f be an s-t flow in D subject to u, and let $D_f = (V, A_f, \psi_f)$ be a residual digraph of f with residual capacity function u_f. Suppose D_f has no s-t paths. Define U as the set of vertices reachable from s in D_f. Then $\mathrm{value}(f) = \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}(U)} u$; consequently, f has maximum value, and $\delta^{\mathrm{out}}(U)$ has minimum capacity.

Proof. We have $t \notin U$ since D_f has no s-t paths. Then, by Corollary 4.5, we have $\mathrm{value}(f) = \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}(U)} f - \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{in}}(U)} f$. If $a \in \delta^{\mathrm{out}}(U)$, then $f(a) = u(a)$; otherwise, $(a, +1) \in A_f$ and U would be extended to contain the head of arc a. If $a \in \delta^{\mathrm{in}}(U)$, then $f(a) = 0$; otherwise, $(a, -1) \in A_f$ and U would be extended to contain the tail of arc a. Therefore, we have $\mathrm{value}(f) = \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}(U)} u$, and by Corollary 4.6, f has maximum value, and $\delta^{\mathrm{out}}(U)$ has minimum capacity. ☐

Theorem 4.10 (max-flow min-cut theorem)**.** Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Then the maximum value of an s-t flow subject to u is equal to the minimum capacity of an s-t cut.

Proof. Let f be an s-t flow in D of maximum value as Proposition 4.7 proves to exist. We will show an s-t cut in D with capacity equals the value of f, and then, by Corollary 4.6, this must be an s-t cut in D of minimum capacity.

Let D_f be the residual digraph of f with residual capacity function u_f. Suppose D_f has an s-t path P. Then, by Proposition 4.8, for $\varepsilon \coloneqq \min\{\, u_f(a) : a \in A(P)\} > 0$, we have that $f' \coloneqq f + \varepsilon \cdot \mathrm{dir}_P$ is an s-t flow in D subject to u with $\mathrm{value}(f') = \mathrm{value}(f) + \varepsilon$, which contradicts the hypothesis about f.

Thus, D_f has no s-t paths. Define U as the set of vertices reachable from s in D_f. Then, by Proposition 4.9, $\mathrm{value}(f) = \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}(U)} u$. ☐

4.2 The Edmonds-Karp Algorithm

From the proof of Theorem 4.10, we now develop the description of an algorithm that solves the maximum flow problem (4.5) efficiently. First, we present the **Ford-Fulkerson method** from which we define the algorithm. We call it a "method" rather than an "algorithm" because it has different implementations with different running times; the algorithm that we present is one implementation of this method.

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. The method starts with the s-t flow $f \coloneqq 0$ in D. While the residual digraph D_f of f has s-t paths, the method does the following

$$\begin{aligned}&\text{Choose an } s\text{-}t \text{ path } P \text{ in } D_f, \text{ and reset } f \text{ to a new } s\text{-}t \text{ flow } f' \coloneqq f + \varepsilon \cdot \mathrm{dir}_P\\ &\text{for } \varepsilon \coloneqq \min\{\, u_f(a) : a \in A(P)\} > 0\,.\end{aligned} \tag{4.10}$$

When D_f has no s-t paths, the s-t flow f is a maximum flow by Proposition 4.9, and the method terminates with f as its output.

The path P in (4.10) is called a **(flow)-augmenting path** as "by pushing flow through it", that is, by resetting f to f', the function f continues to be an s-t flow in D but with greater value than it has before. At a first glance, that seems a promising method to solve the maximum flow problem. However, if in step (4.10) an arbitrary s-t path is chosen, giving rise to the so-called **Ford-Fulkerson algorithm**, some problems arise. First, if the capacity function is allowed to have irrational entries, an instance may be built in which the algorithm does not terminate (see Schrijver [15, Section 10.4a.]). Second, the running time of the so-designed algorithm is not bounded by a polynomial in the input size, that is, in $|V|$, $|A|$, and the size of the representation of the capacities of vector u. Instead, one can show its running time is $O(|A| \cdot C)$ where $C \coloneqq \sum_{a \in \delta_D^{\mathrm{out}}(s)} u_a$, i.e., it is bounded by a polynomial in the magnitude of the capacities numbers and not in the size of the representation of them. Finally, as a matter of style, the algorithm does not have a combinatorial evolution, i.e., the parameters that may characterize the algorithm progress may change by noninteger numbers. For instance, in (4.10), since ε can be any real number, the s-t flow f', and so the resultant f, can have noninteger entries.

Therefore, we show the algorithm due to Edmonds and Karp that makes a subtle but crucial choice — in (4.10), it chooses P as a shortest s-t path of D_f. As we shall see, this algorithm that we call the **Edmonds-Karp algorithm** has all the three properties previously mentioned that the Ford-Fulkerson algorithm lacks.

We start with an auxiliary result. Consider the context of (4.10). If there is an arc that is not in A_f, but it is in $A_{f'}$, then its "reverse" arc is in the augmenting path P.

Proposition 4.11. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Let $f\colon A \to \mathbb{R}$ be an s-t flow subject to u. Let D_f be the residual digraph of f, and let u_f be its residual capacity function. Moreover, let P be a shortest s-t path in the residual digraph D_f, and set $f' := f + \varepsilon \cdot \mathrm{dir}_P$ for $\varepsilon := \min\{\, u_f(a) : a \in A(P)\} > 0$. Also, let $a \in A$ and $\alpha \in \{-1, +1\}$. If $(a, \alpha) \notin A_f$ and $(a, \alpha) \in A_{f'}$, then $(a, -\alpha) \in A(P)$.

Proof. If $\alpha = 1$, then $f(a) = u(a)$ and $f'(a) < u(a)$; thus, $(a, -1) \in A(P)$. Otherwise, $f(a) = 0$ and $f'(a) > 0$; therefore, $(a, +1) \in A(P)$. □

Given a digraph D and any two vertices $u, v \in V(D)$, we define $\mathrm{dist}_D(u, v)$ to be the length of a shortest path between u and v in D. This function will be the first of two parameters that we will use to analyze the Edmonds-Karp algorithm.

Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. As we have described in the beginning of the section, given D, s, t, and u as input, the Ford-Fulkerson method will produce a sequence of s-t flows $f = (f_0 := 0, f_1, \ldots, f_k)$, for some $k \in \mathbb{N}$, where, for each $i \in [k]$, the s-t flow f_i is defined from f_{i-1} as f' is defined from f in (4.10). Next, we show that the function dist is, in a sense given more precisely in Lemma 4.12, *nondecreasing* when we consider, in this order, any two consecutive residual digraphs $D_{f_{i-1}}$ and D_{f_i}, for $i \in [k]$, corresponding to the consecutive s-t flows f_{i-1} and f_i in f. This will imply that dist is "nondecreasing" when we consider, in this order, any two residual digraphs D_{f_i} and D_{f_j}, for $i, j \in [k]$ with $i < j$. This, together with the fact that any distance in D is at most $|V| - 1$, starts to reveal how the algorithm evolves with respect to the function dist. We will get a better grasp on this relation between algorithm progress and dist function in Corollaries 4.14 and 4.15.

Lemma 4.12. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Let $f\colon A \to \mathbb{R}$ be an s-t flow subject to u. Let D_f be the residual digraph of f, and let u_f be its residual capacity function. Moreover, let P be a shortest s-t path in the residual digraph D_f, and set $f' := f + \varepsilon \cdot \mathrm{dir}_P$ for $\varepsilon := \min\{\, u_f(a) : a \in A(P)\} > 0$. Then, for each $w \in V$,

$$\mathrm{dist}_{D_f}(s, w) \le \mathrm{dist}_{D_{f'}}(s, w) \quad \text{and} \quad \mathrm{dist}_{D_f}(w, t) \le \mathrm{dist}_{D_{f'}}(w, t)\,. \tag{4.11}$$

Proof. Suppose that

$$\mathrm{dist}_{D_f}(s, w) > \mathrm{dist}_{D_{f'}}(s, w)\,, \tag{4.12}$$

for some $w \in V$. Let $w \in V$ be such vertex for which $\mathrm{dist}_{D_{f'}}(s, w)$ is minimum. Note that $w \neq s$. Let P' be a shortest path from s to w in $D_{f'}$, and let $v \in V$ be the previous vertex to w in P'. So there is an arc $a \in A$ and a scalar $\alpha \in \{-1, 1\}$ such that $\psi_{f'}((a, \alpha)) = \psi(a)^\alpha = vw$. By the choice of w, we know that

$$\mathrm{dist}_{D_f}(s, v) \le \mathrm{dist}_{D_{f'}}(s, v)\,, \tag{4.13}$$

otherwise, since $\mathrm{dist}_{D_{f'}}(s, v) = \mathrm{dist}_{D_{f'}}(s, w) - 1 < \mathrm{dist}_{D_{f'}}(s, w)$, we would have chosen v in w's place. Thus,

$$\begin{aligned}\mathrm{dist}_{D_f}(s, v) &\le \mathrm{dist}_{D_{f'}}(s, v)\,, \\ &= \mathrm{dist}_{D_{f'}}(s, w) - 1\,, &&\text{for } v \text{ is the previous vertex to } w \text{ in } P'\,, \\ &< \mathrm{dist}_{D_f}(s, w) - 1\,, &&\text{by (4.12)}\,.\end{aligned} \tag{4.14}$$

Then $vw \notin \psi_f(A_f)$, and as ψ_f and $\psi_{f'}$ are defined as in (4.7) and $\psi_{f'}((a,\alpha)) = vw$, we cannot have (a,α) in A_f. However, we know $(a,\alpha) \in A_{f'}$ with $\psi_{f'}((a,\alpha)) = vw$. Thus, by Proposition 4.11, $(a,-\alpha) \in A(P)$, and then $\psi_f((a,-\alpha)) = \psi(a)^{-\alpha} = wv$, i.e., w is the previous vertex to v in P. Then $\operatorname{dist}_{D_f}(s,v) = \operatorname{dist}_{D_f}(s,w)+1$ which with (4.14) leads to $\operatorname{dist}_{D_f}(s,w) + 1 < \operatorname{dist}_{D_f}(s,w) - 1$, a contradiction. Hence, such vertex w does not exist, and the first inequality in (4.11) is proved.

Now consider the reverse digraph D^{-1} of D (see reverse digraph definition in Section 2.3). Set $g \coloneqq f$. Note that g is an t-s flow in D^{-1} subject to u. Let D_g be the residual digraph of g, and let u_g be the residual capacity function of D_g. Note that $D_g = D_f^{-1}$ because D_g is the residual digraph of D^{-1} with respect to $g = f$ and u; also, as $A_g = A_f$ and $g = f$, we have $u_g = u_f$ (see residual capacity function definition in (4.8)). Moreover, note that P^{-1} is a shortest t-s path in D_g since P is a shortest s-t path in D_f and $D_g = D_f^{-1}$. Still, note that $\varepsilon = \min\{\, u_g(a) : a \in A(P^{-1})\}$ and $\operatorname{dir}_P = \operatorname{dir}_{P^{-1}}$ (see definition of vector dir in (4.9)), and set $g' \coloneqq g + \varepsilon \cdot \operatorname{dir}_P$. Finally, note that

$$\operatorname{dist}_{D_f}(w,t) = \operatorname{dist}_{D_g}(t,w) \quad \text{and} \quad \operatorname{dist}_{D_{f'}}(w,t) = \operatorname{dist}_{D_{g'}}(t,w)\,, \tag{4.15}$$

for each $w \in V$. Moreover, by replacing f, f', and s by g, g', and t in the first inequality in (4.11), we have

$$\operatorname{dist}_{D_g}(t,w) \le \operatorname{dist}_{D_{g'}}(t,w) \qquad \text{for each } w \in V\,,$$

Thus, it follows the second inequality in (4.11). □

We introduce the second parameter that we will use to analyze the Edmonds-Karp algorithm. Given a digraph D, define

$$\mu(D) \coloneqq \{\, a \in A : a \in A(P) \text{ for some shortest } s\text{-}t \text{ path } P \text{ in } D\}\,. \tag{4.16}$$

We will see later how this parameter relates with the dist function. But first, we present an important auxiliary result. To present this result and Corollary 4.14, we will build a new digraph D' from D. The idea is to add in D the "reverse" arc of each arc in $\mu(D)$.

Let $D = (V, A, \psi)$ be a digraph. Define the digraph $D^+ = (V, A \times \{-1,+1\}, \psi_+)$ where

$$\psi_+(a,\alpha) \coloneqq \psi(a)^\alpha \qquad \text{for each } (a,\alpha) \in A \times \{-1,+1\}\,. \tag{4.17}$$

Identify each arc $a \in A$ with the arc $(a,+1)$ of D^+ so that we can consider D as a spanning subdigraph of D^+. Define

$$\mu(D)^{-1} \coloneqq \{\, (a,-1) \in A \times \{-1\} : (a,+1) \in \mu(D)\}\,. \tag{4.18}$$

Then define the digraph $D' = (V, A', \psi')$ where

$$A' = A \cup \mu(D)^{-1} \quad \text{and} \quad \psi' = \psi_+{\restriction_{A'}}\,. \tag{4.19}$$

Theorem 4.13. Let $D = (V, A, \psi)$ be a digraph, and let $s, t \in V$ be distinct. Let $D^+ = (V, A\times \{-1,+1\}, \psi_+)$ and $D' = (V, A', \psi')$ be the digraphs defined as in (4.17) and (4.19), respectively. Then $\operatorname{dist}_D(s,t) = \operatorname{dist}_{D'}(s,t)$ and $\mu(D) = \mu(D')$.

Proof. We claim it suffices to show that

$$\operatorname{dist}_D(s,t) \text{ and } \mu(D) \text{ do not change when we add an arc of } \mu(D)^{-1} \text{ in } A\,. \tag{4.20}$$

Suppose, without loss of generality, that $\mu(D) = \{(d_1,+1),(d_2,+1),\ldots,(d_k,+1)\}$ for some $k \in \mathbb{N}$. Consider the sequence of digraphs $(H_0 \coloneqq D, H_1, H_2, \ldots, H_k)$ where, for each $i \in [k]$,

$$H_i \coloneqq (V, B_i, \phi_i)\,, \qquad \text{where } B_i \coloneqq A(H_{i-1}) \cup \{(d_i,-1)\} \text{ and } \phi_i \coloneqq \psi_+{\restriction_{B_i}}\,.$$

Suppose we have shown (4.20). The argument follows by induction on $i \in [k]$. Immediately we would have $\operatorname{dist}_D(s,t) = \operatorname{dist}_{H_1}(s,t)$ and $\mu(D) = \mu(H_1)$. Suppose that, for $i \in [k-1]$, we have $\operatorname{dist}_D(s,t) = \operatorname{dist}_{H_i}(s,t)$ and $\mu(D) = \mu(H_i)$. Moreover, by (4.20) again, we have $\operatorname{dist}_{H_i}(s,t) = \operatorname{dist}_{H_{i+1}}(s,t)$ and $\mu(H_i) = \mu(H_{i+1})$. Thus, $\operatorname{dist}_D(s,t) = \operatorname{dist}_{H_{i+1}}(s,t)$ and $\mu(D) = \mu(H_{i+1})$. This completes the induction, and so we would have the desired since $H_k = D'$. So (4.20) is indeed a sufficient condition for the proof of this theorem.

Now let $(a, +1) \in \mu(D)$ be an arc such that $\psi(a, +1) = uv$ for $u, v \in V$. Also, define the digraph $H \coloneqq (V, B, \phi)$ where $B \coloneqq A \cup \{(a, -1)\}$ and $\phi = \psi_+{\restriction_B}$. Suppose $\mathrm{dist}_D(s,t) \neq \mathrm{dist}_H(s,t)$ or $\mu(D) \neq \mu(H)$. Note that $\mathrm{dist}_D(s,t) \neq \mathrm{dist}_H(s,t)$ and $\mu(D) = \mu(H)$ is false. Then $\mu(D) \neq \mu(H)$ and the shortest s-t paths in H have length at most $\mathrm{dist}_D(s,t)$. Hence, there is an s-t path in H of length at most $\mathrm{dist}_D(s,t)$ that traverses $(a, -1)$. Let $P \coloneqq (u_0 \coloneqq s, a_1, u_1, \dots, a_i, u_i \coloneqq v, a_{i+1} \coloneqq (a, -1), u_{i+1} \coloneqq u, \dots, a_p, u_p \coloneqq t)$ be a such path in H.

Since $(a, +1) \in \mu(D)$, there is a shortest s-t path in D that traverses $(a, +1)$. Let $Q \coloneqq (v_0 \coloneqq s, b_1, v_1, \dots, b_j, v_j \coloneqq u, b_{j+1} \coloneqq (a, +1), v_{j+1} \coloneqq v, \dots, b_q, v_q \coloneqq t)$ be a such path in D. Now set

$$T \coloneqq (V, B', \phi'), \qquad \text{where } B' \coloneqq (A(P) \cup A(Q)) \setminus \{(a, -1), (a, +1)\} \text{ and } \phi' \coloneqq \psi_+{\restriction_{B'}},$$

to be a digraph. Note that T is a subdigraph of D. Moreover, note that $b_{j+2} \in A(T)$, where $\psi_T(b_{j+2}) = v\, v_{j+2} = u_i v_{j+2}$, and so u_i and v_{j+2} are adjacent in T; also, note that $a_{i+2} \in A(T)$, where $\psi_T(a_{i+2}) = u\, u_{i+2} = v_j u_{i+2}$, and so v_j and u_{i+2} are adjacent in T. Thus, if $i < j$, set $W \coloneqq (u_0, a_1, u_1, \dots, a_i, u_i, b_{j+2}, v_{j+2}, b_{j+3}, \dots, b_q, v_q)$ to be an s-t path in T so that the length of W is

$$i + 1 + (q - (j+2)) = \mathrm{dist}_D(s,t) + (i - j) - 1 < \mathrm{dist}_D(s,t).$$

Otherwise, set $W \coloneqq (v_0, b_1, v_1, \dots, b_j, v_j, a_{i+2}, u_{i+2}, a_{i+3}, \dots, a_p, u_p)$ to be an s-t path in T so that the length of W is

$$j + 1 + (p - (i+2)) \leq \mathrm{dist}_D(s,t) + (j - i) - 1 < \mathrm{dist}_D(s,t).$$

In both cases, W is an s-t path in T — a subdigraph of D — of length smaller than the length of a shortest s-t path in D, a contradiction. $\square$

Now we show how $\mu(D)$ relates with $\mathrm{dist}_D(s,t)$. Note that both dist and μ change by integer values, so together they form a combinatorial description of the evolution of the algorithm. Moreover, the way in which both parameters change, as the algorithm evolves, form a lexicographic order that we now define.

The **strict lexicographic order** on the Cartesian product $\mathbb{R} \times \mathbb{R}$ is the binary relation $<^{\mathrm{lex}}$ on $\mathbb{R}$ defined by

$$(a, b) <^{\mathrm{lex}} (a', b') \quad \text{if} \quad a < a' \text{ or } (a = a' \text{ and } b < b') \tag{4.21}$$

for each $a, a', b, b' \in \mathbb{R}$. One can show that $<^{\mathrm{lex}}$ is a *strict total order* (see items (i) to (iii) in Section 2.2), i.e., $<^{\mathrm{lex}}$ is asymmetric, transitive, and semiconnex.

Corollary 4.14. Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Let $f\colon A \to \mathbb{R}$ be an s-t flow subject to u. Let D_f be the residual digraph of f, and let u_f be its residual capacity function. Moreover, let P be a shortest s-t path in the residual digraph D_f, and set $f' \coloneqq f + \varepsilon \cdot \mathrm{dir}_P$ for $\varepsilon \coloneqq \min\{\, u_f(a) : a \in A(P)\} > 0$. Then

(i) if $\mathrm{dist}_{D_{f'}}(s,t) = \mathrm{dist}_{D_f}(s,t)$, then $\mu(D_{f'}) \subsetneq \mu(D_f)$,

(ii) $(\mathrm{dist}_{D_f}(s,t), -|\mu(D_f)|) <^{\mathrm{lex}} (\mathrm{dist}_{D_{f'}}(s,t), -|\mu(D_{f'})|)$.

Proof. **(i)** First, we claim $D_{f'}$ is a subdigraph of $(D_f)'$. Indeed, note that

$$(A_{f'} \setminus A_f) \subseteq A(P)^{-1} \subseteq \mu(D_f)^{-1} \quad \text{and} \quad A((D_f)') = A_f \cup \mu(D_f)^{-1},$$

whence we have $A_{f'} \subseteq A((D_f)')$. Then $\mu(D_{f'}) \subseteq \mu((D_f)')$. Moreover, by Theorem 4.13, $\mu((D_f)') = \mu(D_f)$. Thus, $\mu(D_{f'}) \subseteq \mu(D_f)$.

Now by definition of ε, there exists an arc $(a, \alpha) \in A(P)$ such that $\varepsilon = u_f((a, \alpha))$. If $\alpha = 1$, then $\mathrm{dir}_P((a, 1)) = 1$ and $f'(a) = u(a)$, i.e., $(a, 1) \notin A_{f'}$; if $\alpha = -1$, then $\mathrm{dir}_P((a, -1)) = -1$ and $f'(a) = 0$, i.e., $(a, -1) \notin A_{f'}$. Then $(a, \alpha) \notin A_{f'}$, and so $\mu(D_{f'}) \subsetneq \mu(D_f)$.

(ii) By Lemma 4.12, we have $\mathrm{dist}_{D_f}(s,t) \leq \mathrm{dist}_{D_{f'}}(s,t)$. Suppose $\mathrm{dist}_{D_f}(s,t) = \mathrm{dist}_{D_{f'}}(s,t)$; otherwise, we are done. Then $\mu(D_{f'}) \subsetneq \mu(D_f)$ by item (i), and so $-|\mu(D_f)| < -|\mu(D_{f'})|$. $\square$

Finally, besides the combinatorial and lexicographic pattern that characterizes the progress of the Edmonds-Karp algorithm, the algorithm always terminates, i.e., in a finite number of steps the current residual digraph does not have s-t paths. Moreover, the algorithm terminates in at most $2\,|V|\,|A|$ iterations.

Corollary 4.15 (Complexity of Edmonds-Karp Algorithm)**.** Let $D = (V, A, \psi)$ be a digraph, let $s, t \in V$ be distinct, and let $u\colon A \to \mathbb{R}_+$. Set $m := |A|$, and set $n := |V|$. Let $f = (f_0, f_1, \dots)$ be a sequence of s-t flows in D subject to u with the following properties: $f_0 = 0$ and, for each $i \in \mathbb{Z}_+$, if there is an s-t path in the residual digraph D_{f_i}, then $f_{i+1} := f_i + \varepsilon \cdot \mathrm{dir}_P$ for $\varepsilon := \min\{\, u_{f_i}(a) : a \in A(P)\} > 0$ and P a shortest s-t path in D_{f_i}, otherwise, $f_{i+1} = f_i$. Then $f_{2mn+1} = f_{2mn}$.

Proof. Set $I := \{\, i \in \mathbb{Z}_+ : \text{there exists an } s\text{-}t \text{ path in } D_{f_i}\}$. Note that, by construction of the sequence f,

$$\text{if } r \in I, \text{ then } \ell \in I \text{ for each nonnegative integer } \ell < r. \tag{4.22}$$

So suppose $0 \in I$, otherwise, $I = \varnothing$, and we are done. For convenience, for each $i \in \mathbb{Z}_+$, set $\mathrm{dist}_i := \mathrm{dist}_{D_{f_i}}(s, t)$, and set $\mu_i := \mu(D_{f_i})$.

Consider the binary relation $<^{\mathrm{lex}}$ on $\mathbb{R}$ defined as in (4.21). By item (ii) of Corollary 4.14, for each $i \in I$, we have $(\mathrm{dist}_i, -|\mu_i|) <^{\mathrm{lex}} (\mathrm{dist}_{i+1}, -|\mu_{i+1}|)$. Also, as $<^{\mathrm{lex}}$ is a strict total order, it is transitive (see item (ii) in Section 2.2). Thus, for each distinct i, j in I, we have $(\mathrm{dist}_i, -|\mu_i|) \neq (\mathrm{dist}_j, -|\mu_j|)$. Moreover, note that, for each $i \in I$, we have $\mathrm{dist}_i \in [n-1]$ and $|\mu_i| \in [2m]$. Thus,

$$|I| \leq |[n-1] \times [2m]| \leq 2mn\,. \tag{4.23}$$

Therefore, by (4.22) and (4.23), for each integer $r \geq 2mn + 1$, we have $r \notin I$, and so $f_r = f_{2mn}$. $\square$

4.3 Hoffman's Circulation Theorem

Let $D = (V, A, \psi)$ be a digraph. A vector $f \in \mathbb{R}^A$ is a **circulation** in D if

$$\mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{in}}_D(v)} f = \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}_D(v)} f \qquad \text{for each } v \in V\,. \tag{4.24}$$

The following theorem states a characterization for the existence of circulations.

Theorem 4.16 (Hoffman's circulation theorem)**.** Let $D = (V, A, \psi)$ be a digraph, let $\ell\colon A \to \mathbb{R}_+$ and $u\colon A \to \mathbb{R}_+ \cup \{+\infty\}$ such that $\ell \leq u$. Then there exists a circulation $f \in \mathbb{R}^A$ such that $\ell \leq f \leq u$ if and only if, for every $S \subseteq V$, we have

$$\mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{in}}(S)} \ell \leq \mathbb{1}^{\mathsf{T}}_{\delta^{\mathrm{out}}(S)} u. \tag{4.25}$$

Moreover, the theorem remains true if all occurrences of $\mathbb{R}$ are replaced with $\mathbb{Z}$ throughout all the statements.

Chapter 5

Randomized Algorithms and Sampling Spanning Trees

Recall that our main goal is to show we can run ApproxATSP (see Algorithm 3.1) in polynomial time. This chapter aims to prove Theorem 5.30 that guarantees we can perform Line 2 of ApproxATSP in polynomial time.

Let D, x^*, z^*, G_{z^*} be as in Definition 3.1, and set $n \coloneqq |V|$. In essence, we will round the point z^* to a spanning tree T of G_{z^*} such that no cut of G_{z^*} contains many edges of T (bound given by a constant and z^*), and the cost of T is at most two times the cost of x^*. In a bit more detail, the steps will be (not exactly in presentation order): show z^* belongs to the spanning tree polytope of G_{z^*}; determine a randomized polynomial-time algorithm that rounds z^* to a spanning tree of G_{z^*} and whose probability space has some special properties; show that using this algorithm, we can find, in polynomial time and with high probability, a spanning tree of G_{z^*} that is $(\alpha, 2)$-thin for $\alpha \coloneqq 4 \ln n / \ln \ln n$.

Since Line 2 is the randomized part of ApproxATSP, we start introducing in Section 5.1 some basic definitions and results on probability theory in the discrete case. Then in Section 5.2 we show some so-called concentration bounds, mainly a Chernoff bound for the sum of 0-1 negatively correlated random variables that will be crucial to prove we can find the desired tree with high probability. Next in Section 5.3, we present the Randomized Swap Rounding (RSR), a randomized polynomial-time algorithm that rounds a point in the spanning tree polytope of a graph to a spanning tree. We also show the probability space used by RSR has some special properties. In Section 5.4, we show $z^* \in P_{\text{sptree}}(G_{z^*})$ and that algorithm RSR is indeed the one we need, with some properties, to round z^* to a spanning tree of G_{z^*}. In Section 5.5, given a connected graph with no loops and at least two vertices, we provide a polynomial, in the size of its vertex set, upper bound for the number of cuts with weight at most a factor of the minimum weight of a cut. This result, due to Karger, will be decisive to show the spanning tree of G_{z^*} sampled from RSR, with z^* as input, is α-thin with high probability. Finally, in Section 5.6, we gather all the knowledge acquired — concentration bounds, RSR, Karger's result — to prove Theorem 5.30.

5.1 Discrete Probability

A **finite probability space** is an ordered pair $(\Omega, \mathbb{P})$, where Ω is a nonempty finite set called **sample space** whose elements are called **outcomes** or **elementary events**, and $\mathbb{P}$ is a function from $\mathcal{P}(\Omega)$ to $\mathbb{R}$ such that

$$\mathbb{P}(\{\omega\}) \geq 0 \qquad \text{for each } \omega \in \Omega, \tag{5.1}$$

$$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1 \qquad \text{, and} \tag{5.2}$$

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\}) \qquad \text{for each } E \subseteq \Omega; \tag{5.3}$$

function $\mathbb{P}$ is called the **probability function**, each subset E of Ω is called an **event**, and $\mathbb{P}(E)$ is the *probability* of $E \subseteq \Omega$.[1] For simplicity, we write $\mathbb{P}(\omega)$ instead of $\mathbb{P}(\{\omega\})$ for each $\omega \in \Omega$.

As immediate consequences of the definition of probability function in (5.1) to (5.3), we have

$$\mathbb{P}(\varnothing) = 0\,, \tag{5.4}$$
$$\mathbb{P}(\Omega) = 1 \qquad \text{, and} \tag{5.5}$$
$$\mathbb{P}(\omega) \leq 1 \qquad \text{for each } \omega \in \Omega. \tag{5.6}$$

Since events are sets, it is natural to wonder how set operations between events impact the probability function. In the next proposition we analyze the set operations of union and complementation.

Proposition 5.1. Let $(\Omega, \mathbb{P})$ be a finite probability space. Let A, B be events of Ω. Then

(i) (Principle of Inclusion-Exclusion) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(ii) (Union Bound) For any countable family $\{E_i\}_{i \in I}$ of events of Ω, we have $\mathbb{P}(\cup_{i\in I} E_i) \leq \sum_{i \in I} \mathbb{P}(E_i)$.

(iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if A and B are disjoint.

(iv) $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

(v) $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subseteq B$.

You may have been wondering about the set operation of intersection between events. Now we present the conditional probability, and then the independence of events that use such operation. Let $(\Omega, \mathbb{P})$ be a finite probability space, and let A, B be events of Ω. If $\mathbb{P}(B) \neq 0$, we define

$$\mathbb{P}(A \mid B) \coloneqq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{5.7}$$

to be the **conditional probability** of A given B (similarly for $\mathbb{P}(A \mid B)$). Actually, we have just created another probability function with respect the same sample space. One can show that if $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(\cdot \mid B)\colon S \in \mathcal{P}(\Omega) \mapsto \mathbb{P}(S \mid B)$ satisfies (5.1) to (5.3) and $(\Omega, \mathbb{P}(\cdot \mid B))$ is a finite probability space. In this new probability space, the probability of events outside B is zero, the probability of B is one, and the relative magnitudes of the outcomes inside B is preserved, that is, for each $\omega \in B$ we have $\mathbb{P}(\omega \mid B) = \alpha\, \mathbb{P}(\omega)$ where α ends up being $1/\mathbb{P}(B)$ (the value of α is a consequence of $\mathbb{P}(B \mid B) = 1$ and $\mathbb{P}(\omega \mid B) = 0$ for each $\omega \notin B$).

Before proceeding to the independence definition, we present two important relations.

Proposition 5.2 (Multiplication Rule)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let $\{A_i\}_{i\in[n]}$ be a family of $n \in \mathbb{N}$ events in Ω such that $\mathbb{P}(A_j \mid \cap_{1 \leq i < j} A_i) > 0$ for each $1 < j \leq n$. Then

$$\mathbb{P}\big(\cap_{i\in[n]} A_i\big) = \mathbb{P}(A_1)\, \mathbb{P}(A_2 \mid A_1)\, \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid \cap_{1 \leq i < n} A_i)\,. \tag{5.8}$$

Proposition 5.3 (Law of total probability)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let A be an event of Ω, and let $\{B_i\}_{i\in I}$ be a partition of Ω such that $\mathbb{P}(B_i) > 0$ for each $i \in I$. Then

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i\in I} \mathbb{P}(A)\, \mathbb{P}(A \mid B_i)\,. \tag{5.9}$$

[1] We should mention that this is not the general definition of a probability space, in the sense that it only deals with finite sample spaces. We have chosen this probability framework since it completely describes our study case with a lighter notation and simpler treatment. When the sample space is infinite, and one tries to define the probability function as in (5.3) — define the probability of events as the sum of the probability of outcomes — usually the probability of some event ends up becoming infinity, which violates (5.2). One can refer to [11], for instance, to check out the broader approach that solves this issue; we give a glimpse here.

The general definition of a probability space is as an ordered triple $(\Omega, \Sigma, \mathbb{P})$, where Σ is a set of subsets of Ω satisfying some closure properties that make it a σ-algebra of Ω, and the probability function $\mathbb{P}$ is defined a bit differently. The idea is to define the probability only for the events that lie in Σ, which usually are not all subsets of the sample space if Ω is uncountable; also, the closure properties of a σ-algebra guarantee, in essence, the probability function is defined for all subsets of Ω of interest.

In our case, the sample space Ω is always finite, and the Σ is always $\mathcal{P}(\Omega)$. Thus, we decided to omit Σ, and define this particular case — **finite** probability space — as an ordered pair. Still, we say a *discrete* probability space when the sample space is countable (also called countably infinite).

In particular, if B is an event of Ω such that $\mathbb{P}(B) > 0$ and $\mathbb{P}(\overline{B}) > 0$, then

$$\mathbb{P}(A) = \mathbb{P}(A)\,\mathbb{P}(A \mid B) + \mathbb{P}(A)\,\mathbb{P}(A \mid \overline{B})\,. \tag{5.10}$$

Let $(\Omega, \mathbb{P})$ be a finite probability space, and let A, B be events of Ω. We say A and B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Note that if $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$, then the following are equivalent: A and B are independent; $\mathbb{P}(A \mid B) = \mathbb{P}(A)$; $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. One should not confuse independence with disjoint events. For instance, if $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$ and A, B are disjoint, it follows immediately from definition of independence that A and B are not independent.

Recall that the conditional probability is just a probability function in the same original sample space. Thus, we can consider another definition of independence according to the conditional probability function of some event. Let C be an event of Ω such that $\mathbb{P}(C) \neq 0$. Then given event C, the events A and B are **conditionally independent** if

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)\,, \tag{5.11}$$

i.e., events A, B are independent according to $\mathbb{P}(\cdot \mid C)$ (in the finite probability space $(\Omega, \mathbb{P}(\cdot \mid C))$). Equivalently, one can show using the definition of conditional probability and the Proposition 5.2 that given event C, the events A and B are conditionally independent if

$$\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C) \qquad \text{or} \qquad \mathbb{P}(B \mid C) = 0\,, \tag{5.12}$$

i.e., if $\mathbb{P}(B \mid C) \neq 0$ and given C has occurred, the probability of A does not change if we know that B has occurred ((5.11) is also equivalent to (5.12) if we swap A and B, and this equivalent condition has a similar interpretation to the one we give in this paragraph for (5.12)).

When one analyzes a random experiment, it is common to have more interest in some derived information of the event than in the event itself. For instance, in the experiment of flipping a coin a few times, one may be only interested in the probability that a certain number of heads are obtained and ignores which sequences of coin flips form the corresponding event. Thus, as we next see, the random variables and their distributions come in handy. Ultimately, we will be dealing with the probability of events in a probability space. However, the random variables and their distributions allow us to represent events and compute their probabilities more easily (for instance, the above-mentioned coin toss experiment) and even compute probabilities of an underlying probability space whose sample space is unknown.

Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. A function $X \colon \Omega \to \mathbb{R}$ is called a **random variable**[2] (on $\mathcal{P}$). Let $X \colon \Omega \to \mathbb{R}$ be a random variable. Define the function $\mathbb{P}_X : \mathcal{P}(\mathbb{R}) \to [0, 1]$ as

$$\mathbb{P}_X(S) \coloneqq \mathbb{P}(X^{-1}(S)) = \mathbb{P}(\{\, \omega \in \Omega : X(\omega) \in S \}) \qquad \text{for each } S \subseteq \mathbb{R}\,, \tag{5.13}$$

i.e., $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$; this function is called the **(probability) distribution** of X. Although $\mathbb{P}_X$ is another function in $\mathcal{P}$, we will represent it through $\mathbb{P}$ and predicates involving X. For instance, $\mathbb{P}(X \in S) \coloneqq \mathbb{P}_X(S)$ for each $S \subseteq \mathbb{R}$; also, $\mathbb{P}(X = a) \coloneqq \mathbb{P}_X(a) \coloneqq \mathbb{P}_X(\{a\})$ for each $a \in \mathbb{R}$.

So suppose we have again the random experiment of coin tosses. Then, given the probability space $(\Omega, \mathbb{P})$ that represents it, we could define a random variable $X \colon \Omega \to \mathbb{R}$ where $X(\omega)$ would be the number of heads in the outcome $\omega \in \Omega$; thus, given an arbitrary $x \in \mathbb{N}$ we could look at $\mathbb{P}(X = x)$ instead of $\mathbb{P}(\{\, \omega \in \Omega : X(\omega) = x \})$.

Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space, and let $X \colon \Omega \to \mathbb{R}$ be a random variable. The **expected value** or **expectation** of X, denoted by $\mathbb{E}[X]$, is defined as

$$\mathbb{E}[X] \coloneqq \sum_{x \in \operatorname{Im}(X)} x\,\mathbb{P}(X = x)\,. \tag{5.14}$$

[2] If the probability space $\mathcal{P}$ is an ordered triple $(\Omega, \Sigma, \mathbb{P})$ and Σ is a proper subset of $\mathcal{P}(\Omega)$, then to X be a random variable, it also needs to be Σ-measurable, that is, $X^{-1}(a) \in \Sigma$ for each $a \in \mathbb{R}$. Since in our finite case the sigma-algebra is always $\mathcal{P}(\Omega)$ so that we have omitted it, it is sufficient for us to check that X is a function from Ω to $\mathbb{R}$.

Since Ω is finite[3], we have

$$\mathbb{E}[X] = \sum_{x \in \mathrm{Im}(X)} x \sum_{\substack{\omega \in \Omega \text{ s.t.} \\ X(\omega)=x}} \mathbb{P}(\omega) = \sum_{\omega \in \Omega} \sum_{x \in \mathrm{Im}(X)} x \, [X(\omega) = x] \, \mathbb{P}(\omega) = \sum_{\omega \in \Omega} X(\omega) \, \mathbb{P}(\omega) \,. \tag{5.15}$$

Now we find ways to build new random variables from random variables.

$$\begin{aligned}&\text{Let } \mathcal{P} = (\Omega, \mathbb{P}) \text{ be a finite probability space. Let } X \text{ be a random variable on } \mathcal{P}.\\ &\text{Moreover, let } f\colon \mathbb{R} \to \mathbb{R} \text{ be a function. Since } f \circ X \text{ is a function from } \Omega \text{ to } \mathbb{R},\\ &\text{we have that } f \circ X \text{ is a random variable on } \mathcal{P}\,;\end{aligned} \tag{5.16}$$

we denote $f \circ X$ by $f(X)$, and we say $f(X)$ is a function of X. The next proposition shows how to write the distribution and expectation of $f(X)$ in terms of the distribution of X. Moreover, it shows that applying a function to independent random variables preserves their independence.

Proposition 5.4 (Function of single random variable)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X, Y be random variables on $\mathcal{P}$. Moreover, let f, g be functions from $\mathbb{R}$ to $\mathbb{R}$. Then

(i) (Distribution)

$$\mathbb{P}(f(X) = y) = \sum_{\substack{x \in \mathrm{Im}(X) \text{ s.t.} \\ f(x)=y}} \mathbb{P}(X = x) = \sum_{\substack{x \in \mathrm{Im}(X) \text{ s.t.} \\ f(x)=y}} \mathbb{P}(X = x) \qquad \text{for each } y \in \mathrm{Im}(f(X))\,. \tag{5.17}$$

(ii) (Expectation)

$$\mathbb{E}[f(X)] = \sum_{x \in \mathrm{Im}(X)} f(x) \, \mathbb{P}(X = x) \,. \tag{5.18}$$

(iii) If X and Y are independent, then $f(X)$ and $g(Y)$ are independent.

We can also think of a function of multiple random variables.

$$\begin{aligned}&\text{Let } \mathcal{P} = (\Omega, \mathbb{P}) \text{ be a finite probability space. Let } X, Y \text{ be random variables on } \mathcal{P}.\\ &\text{Also, let } f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R} \text{ be a function. Then } f(X,Y)\colon \omega \in \Omega \mapsto f(X(\omega), Y(\omega)) \in \mathbb{R}\\ &\text{is a random variable on } \mathcal{P} \text{ as it is a function from } \Omega \text{ to } \mathbb{R}\,.\end{aligned} \tag{5.19}$$

Similar to a function of a single random variable, we can present the following relations.

Proposition 5.5 (Function of multiple random variables)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be finite probability space. Let X, Y be random variables on $\mathcal{P}$. Moreover, let $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Then

(i) (Distribution)

$$\mathbb{P}(f(X,Y) = z) = \sum_{\substack{x \in \mathrm{Im}(X), \\ y \in \mathrm{Im}(Y) \\ \text{s.t. } f(x,y)=z}} \mathbb{P}(X = x, Y = y) \,. \tag{5.20}$$

(ii) (Expectation)

$$\mathbb{E}[f(X,Y)] = \sum_{\substack{x \in \mathrm{Im}(X), \\ y \in \mathrm{Im}(Y)}} f(x,y) \, \mathbb{P}(X = x, Y = y) \tag{5.21}$$

Proposition 5.6. Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X, Y be random variables on $\mathcal{P}$. Then

[3] The expected value can also be defined if $\mathrm{Im}(X)$ is countable infinite, with an infinite series, or uncountable, with an integral. In such cases, some convergence issues may arise which require a more detailed examination to determine when the expectation indeed exists. In our case, since Ω is finite, we do not have such a problem, and we can directly have the equivalent expression in (5.15) for the expectation of X.

(i) (Constant random variable) If X is a constant, that is, for some $c \in \mathbb{R}$, we have $X(\omega) = c$ for each $\omega \in \Omega$, then $\mathbb{E}(X) = c$.

(ii) (Linearity of expectation) Let $\lambda \in \mathbb{R}$. Then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ and $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$.

(iii) (0-1 random variable) If X is a 0-1 random variable, that is, $\operatorname{Im}(X) = \{0, 1\}$, then $\mathbb{E}[X] = \mathbb{P}(X = 1)$.

(iv) (Product of independent random variables) If X and Y are independent, then $\mathbb{E}[X\,Y] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

(v) (Cauchy-Schwarz inequality)
$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\,\mathbb{E}[Y^2]\,. \tag{5.22}$$

Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X, Y be random variables on $\mathcal{P}$, and let $y \in \mathbb{R}$. Also, set $A \coloneqq \{\, \omega \in \Omega : Y(\omega) = y \}$. Suppose $\mathbb{P}(Y = y) > 0$, and so $\mathbb{P}(A) > 0$. Then the **conditional expectation** of X given A is the scalar defined by
$$\mathbb{E}[X \mid A] \coloneqq \sum_{x \in \operatorname{Im}(X)} x\,\mathbb{P}(X = x \mid A)\,. \tag{5.23}$$
Similarly, the **conditional expectation** of X given that $Y = y$ is defined by
$$\mathbb{E}[X \mid Y = y] \coloneqq \sum_{x \in \operatorname{Im}(X)} x\,\mathbb{P}(X = x \mid Y = y)\,. \tag{5.24}$$
Finally, one can show that for an arbitrary function $f \colon \mathbb{R} \to \mathbb{R}$ it holds
$$\mathbb{E}[f(X) \mid A] \coloneqq \sum_{x \in \operatorname{Im}(X)} f(x)\,\mathbb{P}(X = x \mid A)\,. \tag{5.25}$$

Note that in $\mathbb{E}[X \mid A]$ we can consider that we are dealing with the random variable X on the probability space $(\Omega, \mathbb{P}(\cdot \mid A))$. Thus, (5.23) follows directly from (5.14). Now we present different versions of the so-called *law of total expectation.*

Proposition 5.7 (Law of total expectation)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X, Y, Z be random variables on $\mathcal{P}$. Also, Let $\{A_i\}_{i \in I}$ be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each $i \in I$. Then
$$\mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[X \mid A_i]\,\mathbb{P}(A_i)\,. \tag{5.26}$$
Moreover, for any event B of Ω such that $\mathbb{P}(A_i \cap B) > 0$ for each $i \in I$, we have
$$\mathbb{E}[X \mid B] = \sum_{i \in I} \mathbb{E}[X \mid A_i \cap B]\,\mathbb{P}(A_i \mid B)\,. \tag{5.27}$$
Set $Y^+ \coloneqq \{\, y \in \operatorname{Im}(Y) : \mathbb{P}(Y = y) > 0 \}$. Then
$$\mathbb{E}[X] = \sum_{y \in Y^+} \mathbb{E}[X \mid Y = y]\,\mathbb{P}(Y = y)\,. \tag{5.28}$$
Finally, set $Z^+ \coloneqq \{\, z \in \operatorname{Im}(Z) : \mathbb{P}(Z = z) > 0 \}$. Then for any $z \in Z^+$, we have
$$\mathbb{E}[X \mid Z = z] = \sum_{\substack{y \in Y^+ \text{ s.t.} \\ \mathbb{P}(Y = y\,,\, Z = z) > 0}} \mathbb{E}[X \mid Y = y\,,\, Z = z]\,\mathbb{P}(Y = y \mid Z = z)\,. \tag{5.29}$$

Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X, Y be random variables on $\mathcal{P}$. We can rewrite the law of total expectation if we introduce the following. Define the function
$$\mathbb{E}[X \mid Y] \colon \omega \in \Omega \mapsto \mathbb{E}[X \mid Y = Y(\omega)] \in \mathbb{R}\,. \tag{5.30}$$
Note that $\mathbb{E}[X \mid Y]$ is a function of Y (see (5.16)), and so a random variable on $\mathcal{P}$. Moreover, suppose $\mathbb{P}(Y = y) > 0$ for each $y \in \operatorname{Im}(Y)$. Then, by (5.18),
$$\mathbb{E}[\,\mathbb{E}[X \mid Y]\,] = \sum_{y \in \operatorname{Im}(Y)} \mathbb{E}[X \mid Y = y]\,\mathbb{P}(Y = y) = \mathbb{E}[X]\,, \tag{5.31}$$
i.e., we have an alternative and compact form of writing the law of total of expectation.

5.2 Concentration Bounds

In this section we find bounds for the probability that a random variable deviates from some value, typically its mean. Such bounds are called **concentration bounds**, and the associated probabilities and inequalities are called **tail probabilities** and **concentration inequalities**, respectively.

Markov's inequality presents the first bound. We will use this bound in the proof of Theorem 5.27, and to prove the other concentration bounds of this section. To present this next result we make the following definition. A random variable X on a probability space $(\Omega, \mathbb{P})$ is **nonnegative** if $\mathbb{P}(X < 0) = 0$.

Theorem 5.8 (Markov's inequality)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X be a nonnegative random variable on $\mathcal{P}$. Then, for each $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}\,. \tag{5.32}$$

Proof. Let a be a positive real. For each $x \in \operatorname{Im}(X)$ we have either $x < a$ or $x \geq a$. Set A to be the set $\{\, x \in \operatorname{Im}(X) : x \geq a \}$ so that $\overline{A} := \operatorname{Im}(X) \setminus A = \{\, x \in \operatorname{Im}(X) : x < a \}$. Suppose A is nonempty; otherwise, $\mathbb{P}(X \geq a) = 0$, and so with $\mathbb{E}[X] \geq 0$, since X is nonnegative, we have (5.32). Then

$$\mathbb{E}[X] = \sum_{x \in \overline{A}} x\,\mathbb{P}(X = x) + \sum_{x \in A} x\,\mathbb{P}(X = x) \geq \sum_{x \in A} x\,\mathbb{P}(X = x) \geq a \sum_{x \in A} \mathbb{P}(X = x) = a\,\mathbb{P}(X \geq a)\,, \tag{5.33}$$

where the first inequality holds since X and $\mathbb{P}$ are nonnegative (X is nonnegative by hypothesis and $\mathbb{P}$ is a probability function, so it satisfies (5.1) and (5.3)), and the second inequality holds since $\mathbb{P}$ is nonnegative. This shows (5.32). $\square$

Note that the bound given by Markov's inequality decreases linearly with respect to the deviation term. We can go a bit further with the **Chebyshev's inequality**. This concentration bound decreases quadratically with the deviation term.

Theorem 5.9 (Chebyshev's inequality)**.** Let $\mathcal{P} = (\Omega, \mathbb{P})$ be a finite probability space. Let X be a random variable on $\mathcal{P}$. Then, for each $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\operatorname{Var}(X)}{a^2}\,.$$

Proof. Let a be a positive real. Since $a > 0$, we have $(X - \mathbb{E}[X])^2 \geq a^2$ if and only if $|X - \mathbb{E}[X]| \geq a$. Thus,

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}[X]| \geq a) &= \sum_{\substack{x \in \operatorname{Im}(X) \text{ s.t.} \\ |X - \mathbb{E}[X]| \geq a}} \mathbb{P}(X = x) = \sum_{\substack{x \in \operatorname{Im}(X) \text{ s.t.} \\ (X - \mathbb{E}[X])^2 \geq a^2}} \mathbb{P}(X = x) \\ &= \mathbb{P}\big((X - \mathbb{E}[X])^2 \geq a^2\big)\,.\end{aligned}$$

Then, by Markov's inequality (see Theorem 5.8),

$$\mathbb{P}\big((X - \mathbb{E}[X])^2 \geq a^2\big) \leq \frac{\mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big]}{a^2} = \frac{\operatorname{Var}(X)}{a^2}\,. \qquad \square$$

The proof of Chebyshev's inequality introduces a significant pattern for deriving or proving concentration bounds. That is, for a tail probability one wants to bound, find an equivalent tail probability that is somehow more convenient to apply Markov's inequality.

We will exploit this pattern to find another crucial concentration bound from a family of bounds called **Chernoff bounds**. These bounds can be found using the presented pattern with the equivalent tail probability involving the moment-generating function.

The **moment-generating function** of a random variable X is the function $M_X(t) \colon t \in \mathbb{R} \mapsto \mathbb{E}[e^{tX}]$. This function can be used to find the moments of X. Informally and ignoring issues and details of convergence and differentiation, for instance, we give a glimpse of how one could find a moment of X using such function. From the Taylor expansion of e^{tX} we have

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots + \frac{t^nX^n}{n!} + \cdots . \tag{5.34}$$

Since e^{tX} is a function of a single random variable, by (5.16), e^{tX} is a random variable. Then

$$M_X(t) = \mathbb{E}[e^{tX}] = 1 + t\mathbb{E}[X] + \frac{t^2\mathbb{E}[X^2]}{2!} + \frac{t^3\mathbb{E}[X^3]}{3!} + \cdots + \frac{t^n\mathbb{E}[X^n]}{n!} + \cdots . \tag{5.35}$$

Hence, if one differentiates $i \geq 0$ times the moment-generating function $\mathbb{E}[e^{tX}]$, the resulting function for $t = 0$ will be the ith moment. For instance, $M_X(0)$ is the 0-th moment, the total probability 1, and $M'_X(0)$ is the 1-st moment, the expectation of X.

Now we present a series of four lemmas that we will use to prove a Chernoff bound for the sum of 0-1 random variables that are negatively correlated (we define this property later). The proof of this bound will start with Lemma 5.10 where we will use the discussed pattern with the moment-generating function. Note that in this lemma we do not assume any relation (negative correlation in our case) between the random variables.

Lemma 5.10. Let $(\Omega, \mathbb{P})$ be a finite probability space. Let $X_1, \ldots, X_k \colon \Omega \to \{0, 1\}$ be random variables for some integer $k \geq 1$. Set $X \coloneqq \sum_{i=1}^{k} X_i$. Also, let $\delta > 0$. Then

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) = \mathbb{P}\Big(e^{tX} \geq e^{t(1+\delta)\mathbb{E}[X]}\Big) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mathbb{E}[X]}} \qquad \text{for each } t > 0\,. \tag{5.36}$$

Proof. Let $t > 0$, and let $c \in \mathbb{R}$. Since X is a random variable, $\operatorname{Im}(X)$ is finite, and by the distribution of a function of a random variable (see (5.17) of Proposition 5.4), we have

$$\mathbb{P}(e^{tX} \geq e^{tc}) = \sum_{\substack{x \in \operatorname{Im}(X) \\ \text{s.t. } e^{tx} \geq e^{tc}}} \mathbb{P}(X = x)\,.$$

Also, since $x \in \mathbb{R} \mapsto e^{tx}$ is monotonically increasing, for each $x \in \operatorname{Im}(X)$ we have $e^{tx} \geq e^{tc}$ if and only if $x \geq c$. Then

$$\mathbb{P}(e^{tX} \geq e^{tc}) = \sum_{\substack{x \in \operatorname{Im}(X) \\ \text{s.t. } x \geq c}} \mathbb{P}(X = x) = \mathbb{P}(X \geq c)\,. \tag{5.37}$$

Thus, if we replace the scalar c by the scalar $(1+\delta)\mathbb{E}[X]$ in (5.37), it follows that

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) = \mathbb{P}\Big(e^{tX} \geq e^{t(1+\delta)\mathbb{E}[X]}\Big) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mathbb{E}[X]}}\,, \tag{5.38}$$

where the last inequality follows from Markov's inequality (see Theorem 5.8) since e^{tX} is a nonnegative random variable and $e^x > 0$ for each $x \in \mathbb{R}$. ∎

In Lemma 5.10, for a random variable X that was a sum of 0-1 random variables and a scalar $\delta > 0$, we have presented the tail probability $\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X])$ for which we will show a Chernoff bound in Theorem 5.14, and we have found a family of bounds for it in function of some $t > 0$.

With the next two lemmas, we will bound the $\mathbb{E}[e^{tX}]$ in the RHS of (5.36) with some exponential function involving t and $\mathbb{E}[X]$. Before that, we introduce the crucial definition of this section that we start using in the next lemma.

We say 0-1 random variables $X_1, \ldots, X_n$ are **negatively correlated** if

$$\mathbb{P}(\wedge_{i\in I} X_i = 1) \leq \prod_{i \in I} \mathbb{P}(X_i = 1) \qquad \text{for each } I \subseteq [n] \tag{5.39}$$

or, equivalently,

$$\mathbb{E}\Bigg[\prod_{i\in I} X_i\Bigg] \leq \prod_{i\in I} \mathbb{E}[X_i] \qquad \text{for each } I \subseteq [n]\,. \tag{5.40}$$

Lemma 5.11. Let $(\Omega, \mathbb{P})$ be a finite probability space. Let $X_1, \dots, X_k \colon \Omega \to \{0,1\}$ be random variables, for some integer $k \geq 1$, that are negatively correlated. Then the random variables $e^{tX_1}, \dots, e^{tX_k}$, for some $t > 0$, are also negatively correlated.

Proof. Set $X \coloneqq \sum_{i=1}^k X_i$. Since $\mathbb{E}\big[e^{tX}\big] = \mathbb{E}\Big[\prod_{i=1}^k e^{tX_i}\Big]$, it suffices to show

$$\mathbb{E}\big[e^{tX}\big] \leq \prod_{i=1}^k \mathbb{E}\big[e^{tX_i}\big]\,. \tag{5.41}$$

Define independent 0-1 random variables $Y_1, \dots, Y_k$ that have distribution equal to $X_1, \dots, X_k$, respectively, i.e., $\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i = 1) =: p_i$ for each $i \in [k]$ (note that we do not require the Y_i's to be in the same probability space of the X_i's). Also, set $Y \coloneqq \sum_{i=1}^k Y_i$. By (5.16), e^{tY_i} is a random variable for each $i \in [k]$, and, by (5.19), e^{tY} is a random variable. Moreover, note that, by item (iii) of Proposition 5.4, the e^{tY_i}'s are independent since the Y_i's are independent. Then

$$\mathbb{E}\big[e^{tY}\big] = \mathbb{E}\left[\prod_{i=1}^k e^{tY_i}\right] = \prod_{i=1}^k \mathbb{E}\big[e^{tY_i}\big] = \prod_{i=1}^k \mathbb{E}\big[e^{tX_i}\big]\,,$$

where the second equality holds since the e^{tY_i}'s are independent and by item (iv) of Proposition 5.6. Thus,

$$\text{to prove (5.41), it suffices to show } \mathbb{E}[e^{tX}] \leq \mathbb{E}[e^{tY}]\,. \tag{5.42}$$

Since $t > 0$, by the expansion of the moment-generating function of an arbitraty random variable in (5.35), we have

$$\mathbb{E}[e^{tX}] \leq \mathbb{E}[e^{tY}] \quad \text{if} \quad \mathbb{E}[X^\alpha] \leq \mathbb{E}[Y^\alpha] \quad \text{for each integer } \alpha \geq 1. \tag{5.43}$$

So let $\alpha \geq 1$ be an integer. Note that X^α and Y^α are polynomials whose terms, apart from the coefficients, are all the products of the form $\prod_{i=1}^k X_i^{\alpha_i}$ and $\prod_{i=1}^k Y_i^{\alpha_i}$, respectively, for some integers $\alpha_1, \dots, \alpha_k \geq 0$ with $\sum_{i=1}^k \alpha_i = \alpha$. Hence, by linearity of expectation (see item (ii) of Proposition 5.6),

$$\mathbb{E}[X^\alpha] \leq \mathbb{E}[Y^\alpha] \quad \text{if} \quad \mathbb{E}\left[\prod_{i=1}^k X_i^{\alpha_i}\right] \leq \mathbb{E}\left[\prod_{i=1}^k Y_i^{\alpha_i}\right] \quad \text{for each integer } \alpha_1, \dots, \alpha_k \geq 0 \text{ with } \sum_{i=1}^k \alpha_i = \alpha\,. \tag{5.44}$$

Let $\alpha_1, \dots, \alpha_k \geq 0$ be integers such that $\sum_{i=1}^k \alpha_i = \alpha$. Then, by expected value of a function of multiple random variables (see (5.21) of Proposition 5.5), one has

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^k X_i^{\alpha_i}\right] &= \sum_{x_1 \in \mathrm{Im}(X_1), \dots} \left(\prod_{i=1}^k x_i^{\alpha_i}\right) \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) \\
&= \sum_{x_1 \in \mathrm{Im}(X_1), \dots} \left(\prod_{i=1}^k x_i\right) \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) \quad \text{since } X_i\text{'s are 0-1 random variables} \\
&= \mathbb{E}\left[\prod_{i=1}^k X_i\right] \\
&\leq \prod_{i=1}^k \mathbb{E}[X_i] \qquad \text{since } X_i\text{'s are negatively correlated} \qquad\qquad (5.45) \\
&= \prod_{i=1}^k \mathbb{E}[Y_i] \qquad \text{since } X_i\text{'s and } Y_i\text{'s are both 0-1 r.v.'s with the same distribution} \\
&= \prod_{i=1}^k \mathbb{E}[Y_i^{\alpha_i}] \quad \text{since } Y_i\text{'s are 0-1 random variables and by (5.18) of Proposition 5.4} \\
&= \mathbb{E}\left[\prod_{i=1}^k Y_i^{\alpha_i}\right],
\end{aligned}$$

where the last equality holds as the Y_i's are independent, and so the $Y_i^{\alpha_i}$'s are independent by item (iii) of Proposition 5.4, and by item (iv) of Proposition 5.6.

Therefore, with (5.45), we have shown $\mathbb{E}[X^\alpha] \le \mathbb{E}[Y^\alpha]$ by (5.44), which in turn shows $\mathbb{E}[e^{tX}] \le \mathbb{E}[e^{tY}]$ by (5.43), which finally shows (5.41) by (5.42), and then completes the proof. $\square$

Lemma 5.12. Let $(\Omega, \mathbb{P})$ be a finite probability space. Let $X_1, \dots, X_k \colon \Omega \to \{0,1\}$ be random variables for some integer $k \ge 1$. Set $X \coloneqq \sum_{i=1}^k X_i$. Also, let $t > 0$. Then

$$\prod_{i=1}^{k} \mathbb{E}\big[e^{tX_i}\big] \le e^{(e^t - 1)\mathbb{E}[X]}\,.$$

Proof. Set $p_i \coloneqq \mathbb{P}(X_i = 1)$ for each $i \in [k]$. Since $X_1, \dots, X_k$ are 0-1 random variables, for any $i \in [k]$ the random variable e^{tX_i} is either e^t with probability p_i or 1 with probability $1 - p_i$. Thus,

$$\begin{aligned}\prod_{i=1}^{k} \mathbb{E}\big[e^{tX_i}\big] &= \prod_{i=1}^{k}(1 - p_i + p_i e^t) = \prod_{i=1}^{k}(1 + p_i(e^t - 1)) \le \prod_{i=1}^{k} e^{p_i(e^t-1)} \\ &= e^{(e^t-1)\left(\sum_{i=1}^k p_i\right)} = e^{(e^t-1)\left(\sum_{i=1}^k \mathbb{E}[X_i]\right)} = e^{(e^t-1)\mathbb{E}[X]}\,,\end{aligned}$$

where the first inequality holds since $e^x \ge 1 + x$ for each $x \in \mathbb{R}$ (see Proposition 2.3), and the second last equality holds since X_i's are 0-1 random variables, and so $p_i = \mathbb{E}(X_i)$. $\square$

As we will see in the proof of Theorem 5.14, the last three lemmas will produce a family of bounds in function of some scalar $t > 0$ that have the form of an exponential function involving t itself, a scalar $\delta > 0$, and the expected value of a random variable X that is a sum of 0-1 random variables that are negatively correlated. In the next lemma, we abstract this exponential function and find which point minimizes it.

Lemma 5.13. Let $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\alpha_1 > 0$ and $\alpha_2 \ne 0$. Let $f \colon x \in \mathbb{R} \mapsto e^{(e^x - 1 - \alpha_1 x)\alpha_2}$ be a function. Then $\ln(\alpha_1)$ is a minimum point of f.

Proof. One can show that any critical point of this function is a minimizer. Recall that a critical point is one where the first derivative of the function is zero. Thus, we take the first derivative of f and set it to zero, that is,

$$\begin{aligned}\left(e^{(e^x-1-\alpha_1 x)\alpha_2}\right)' = 0 &\Leftrightarrow e^{((e^x-1-\alpha_1 x)\alpha_2)}((e^x - 1 - \alpha_1 x)\alpha_2)' = 0 \\ &\Leftrightarrow ((e^x - 1 - \alpha_1 x)\alpha_2)' = 0 && \text{since } e^x \ne 0 \text{ for any } x \in \mathbb{R} \\ &\Leftrightarrow (e^x - \alpha_1)\alpha_2 = 0 \\ &\Leftrightarrow x = \ln(\alpha_1) && \text{since } \alpha_1 > 0 \text{ and } \alpha_2 \ne 0\,.\end{aligned}$$

$\square$

Now we join the four last lemmas to find a Chernoff bound for the sum of 0-1 random variables that are negatively correlated. We will use it to prove Lemma 5.25 — a bound regarding the probability that an arbitrarily chosen cut violates the α-thinness property.

Note that independence between random variables implies negative correlation between the same random variables. Hence, this Chernoff bound can be applied for sums of 0-1 random variables that are independent.

Also, note that this Chernoff bound decreases exponentially in the mean and the deviation δ, while, as we have seen, the Markov and Chebyshev inequalities decrease linearly and quadratically, respectively. Thus, for a specific random variable that we see next, this Chernoff bound represents a significant improvement when compared with the inequalities of Markov and Chebyshev.

Theorem 5.14 (Chernoff Bound for sum of 0-1 random variables that are negatively correlated)**.** Let $(\Omega, \mathbb{P})$ be a finite probability space. Let $X_1, \dots, X_k \colon \Omega \to \{0,1\}$ be random variables, for some integer $k \ge 1$, that are negatively correlated. Also, set $X \coloneqq \sum_{i=1}^k X_i$. Then for every $\delta > 0$ we have

$$\mathbb{P}(X \ge (1+\delta)\mathbb{E}[X]) \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}. \tag{5.46}$$

Proof. First, by Lemma 5.10 we have

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) = \mathbb{P}\Big(e^{tX} \geq e^{t(1+\delta)\mathbb{E}[X]}\Big) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mathbb{E}[X]}} \qquad \text{for each } t > 0\,. \tag{5.47}$$

Since $X_1, \ldots, X_k$ are negatively correlated, we have that $e^{tX_1}, \ldots, e^{tX_k}$ are negatively correlated by Lemma 5.11, i.e.,

$$\mathbb{E}\left[\prod_{i=1}^{k} e^{tX_i}\right] \leq \prod_{i=1}^{k} \mathbb{E}\big[e^{tX_i}\big]\,.$$

Thus, as $\mathbb{E}[e^{tX}] = \mathbb{E}\Big[\prod_{i=1}^{k} e^{tX_i}\Big]$, we have

$$\mathbb{E}[e^{tX}] \leq \prod_{i=1}^{k} \mathbb{E}[e^{tX_i}] \leq e^{(e^t-1)\mathbb{E}[X]}\,, \tag{5.48}$$

where the last inequality holds by Lemma 5.12. Therefore, by (5.47) and (5.48),

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq \frac{e^{(e^t-1)\mathbb{E}[X]}}{e^{t(1+\delta)\mathbb{E}[X]}} = e^{(e^t-1-t(1+\delta))\mathbb{E}[X]} \qquad \text{for each } t > 0. \tag{5.49}$$

As we are interested in the lowest upper bound, we look for a $t > 0$ that minimizes this last function. By Lemma 5.13, such point is $\ln(1+\delta)$. Thus, by replacing t by $\ln(1+\delta)$ in the RHS of equation in (5.49), we have

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq \left(\frac{e^{e^{\ln(1+\delta)}-1}}{e^{(1+\delta)\ln(1+\delta)}}\right)^{\mathbb{E}[X]} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mathbb{E}[X]}. \qquad \square$$

5.3 Randomized Swap Rounding

In this section, we present a polynomial-time randomized algorithm that rounds a point in the spanning tree polytope of a graph to the incidence vector of a spanning tree of the graph. Moreover, we show the probability space used by this randomized algorithm has some properties that will allow us to sample an α-thin tree, for a desired value of α, with high probability. Thus, this rounding algorithm is crucial for the development of our main algorithm, the ApproxATSP (see Algorithm 3.1) for the (mATSP); in particular, we use it in Line 2 of that algorithm.

The algorithm is called **randomized swap rounding**, or RSR, and it is due to Chekuri, Vondrak, and Zenklusen [5]. Actually, the algorithm described in that paper shows how to perform rounding in a more general setting, namely rounding a point in a matroid polytope to the incidence vector of an independent set of the matroid (see [16, Chapter 39]). The main routine of the algorithm is SwapRound(G, x), for a graph $G = (V, A, \psi)$ and a point x in its spanning tree polytope, that successively calls Merge$(G, \beta_1, T_1, \beta_2, T_2)$ for some scalars $\beta_1, \beta_2 \in [0, 1]$ and spanning trees T_1, T_2 of G. Note that in the next algorithm and in some proofs that follow, there will be operations of addition and subtraction involving graphs, subgraphs, vertex and edge sets. These are defined in (2.9). Also, for a graph G, the set of spanning trees of G is denoted by $\mathcal{T}(G)$.

Algorithm 5.1: SwapRound(G, x)

Input:
(i) a graph $G = (V, A, \psi)$,
(ii) a point x in $P_{\text{sptree}}(G)$.

Output: A random spanning tree T of G, where for each $e \in E$ the associated 0-1 random variable $X_e := [e \in T]$ has expectation equal to x_e, and the random variables in $\{X_e\}_{e \in E}$ are negatively correlated (see (5.40)).

1. Find $\lambda_1, \ldots, \lambda_r \geq 0$ with $\sum_{\ell=1}^{r} \lambda_\ell = 1$ and $T_1, \ldots, T_r \in \mathcal{T}(G)$ such that $x = \sum_{\ell=1}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$
2. $T \leftarrow T_1$
3. **for** $k = 1$ **to** $r - 1$ **do**
4. $\quad T \leftarrow \text{Merge}(\sum_{\ell=1}^{k} \lambda_\ell, T, \lambda_{k+1}, T_{k+1})$
5. **return** T

Algorithm 5.2: Merge$(G, \beta_1, T_1, \beta_2, T_2)$

Input:
(i) a graph $G = (V, A, \psi)$,
(ii) scalars $\beta_1, \beta_2 \in \mathbb{R}_{++}$,
(iii) spanning trees T_1, T_2 of G.

Output: A spanning tree of G.

1. **while** $T_1 \neq T_2$ **do**
2. $\quad$ Pick $i \in E(T_1) \setminus E(T_2)$ and find $j \in E(T_2) \setminus E(T_1)$ s.t. $T_1 - i + j$ and $T_2 - j + i$ lie in $\mathcal{T}(G)$
3. $\quad$ With probability $\frac{\beta_1}{\beta_1 + \beta_2}$, set $T_2 \leftarrow T_2 - j + i$. Else set $T_1 \leftarrow T_1 - i + j$
4. **return** T_1

Before we look at the complexity of RSR in Theorem 5.16, we analyze Line 2. It involves an exchange of edges between two spanning trees that produces new spanning trees of the graph. Moreover, this exchange property actually works in a more general setting that involves any two bases of a matroid.

Proposition 5.15 (Exchange Property,[16, Theorem 39.12.])**.** Let $G = (V, E, \psi)$ be a graph. Let T_1, T_2 be distinct spanning trees of G, and let $i \in E(T_1) \setminus E(T_2)$. Then there exists an edge $j \in E(T_2) \setminus E(T_1)$ such that $T_1 - i + j$, $T_2 - j + i \in \mathcal{T}(G)$. Moreover, such edges can be found in time polynomial in $|V| + |E|$.

Proof. Set u and v to be the ends of edge i. Note that $T_2 + i$ has a unique circuit C; otherwise, T_2 would have more than one path between u and v, i.e., T_2 would have a circuit. Also, note that $T_1 - i$ is a graph with two components, one containing vertex u and another containing vertex v. Moveover, $P := C - i$ is a path between u and v. Thus, $G' := T_1 - i + P$ is a connected spanning subgraph of G (see definitions in (2.9)).

Now we show there exists an edge $j \in E(P)$, and so an edge of T_2, such that $T_1 - i + j$ is a spanning tree of G. Set S to be the set of acyclic subgraphs of G', and consider the poset $(S, \subseteq)$; see definition in Section 2.2. We claim the following.

$$\text{Let } H \in S. \text{ Then } H \text{ is maximal if and only if } H \text{ is connected and } V(H) = V, \text{ i.e., } H \text{ is a spanning tree of } G' \text{ and } G. \tag{5.50}$$

We show necessity by contrapositive. First suppose H is disconnected. Then there exists an edge $e \in E(G')$ such that $H + e \in S$. Now suppose $V(H) \neq V$. Since G' is connected, there exists an edge $e \in E(G')$ with one end in $V(H)$ and another end in $V(G') \setminus V(H)$, and so $H + e \in S$.

Now we show sufficiency. Let $e \in E(G') \setminus E(H)$. Since $V(H) = V(G')$, the ends of edge e lie in $V(H)$. Then, since H is connected, there exists a path in H between the ends of e. Hence, graph $H + e$ has a circuit. This proves (5.50).

Thus, since the graph $T_1 - i$ is disconnected and belongs to S, there exists $j \in E(P)$ such that $T_1 - i + j$ is acyclic. Moreover, as $T_1 - i$ has two components, graph $T_1 - i + j$ is connected, and so it is a spanning tree of G' and G.

Now, since T_2 is a spanning tree and C is the unique circuit of $T_2 + i$ with $i, j \in E(C)$, by removing edge j from the graph $T_2 + i$, it becomes acyclic but still connected and with vertex set equal to V, i.e., the graph $T_2 + i - j = T_2 - j + i$ (see definition in (2.9) to see the equality) is a spanning tree of G.

Finally, to find edge i we just check for each edge of T_1 whether it is an edge of T_2. To find edge j, we first determine the circuit C using, for instance, a depth-first search, dfs for short. Then for each edge j of the path $P = C - i$, we can test whether $T_1 - i + j$ is a spanning tree of G using a dfs again. $\square$

Theorem 5.16 (Complexity of Randomized Swap Rounding)**.** Given a point $x \in P_{\text{sptree}}(G)$ for a graph $G = (V, E, \psi)$, the randomized swap rounding algorithm runs in polynomial time.

Proof. By [16, Theorem 51.5.], a representation of x as a convex combination, as in Line 1, can be found in polynomial time. So the number of terms of such representation is also polynomial, and we denote it by r.

Then the algorithm calls $r - 1$ times Merge (see Algorithm 5.2) that, apart from scalars, receives two spanning trees T_1 and T_2 of G. Its loop in Line 1 ends when T_1 and T_2 are equal, or equivalently, when $|E(T_1) \cap E(T_2)| = |V| - 1$. Also, note that at each iteration of this loop, $|E(T_1) \cap E(T_2)|$ increases exactly by one. Hence, the total number of iterations of loop in Line 1 is upper bounded by $|V|$.

Finally, by Proposition 5.15, edges i, j such as in Line 2 always exist and can be found in polynomial time. $\square$

By Theorem 5.16, we know that the scalar r in Line 1 of Algorithm 5.1 is bounded by a polynomial of the input size of SwapRound. Now we give an explicit upper bound. Let $G = (V, E, \psi)$ be a graph, let $x \in P_{\text{sptree}}(G)$, and consider a call to SwapRound with input G and x.

On the one hand, by definition in (3.4), $P_{\text{sptree}}(G)$ is the convex hull of incidence vectors of the spanning trees of G. Moreover, we have Carathеódory's theorem that we present without proof.

Theorem 5.17 (Caratheódory's theorem, [17, Corollary 7.1j])**.** Let V be a finite set, and let $X \subseteq \mathbb{R}^V$. If $x \in \operatorname{conv} X$, then there exist affinely independent vectors $x_0, \ldots, x_d \in X$ such that $x \in \operatorname{conv}\{x_0, \ldots, x_d\}$. In particular, $d \leq \dim(\operatorname{conv} X)$.

Thus, point x can be written as a convex combination of at most $\dim(P_{\text{sptree}}(G)) + 1$ terms. On the other hand, recall that $P_{\text{sptree}}(G)$ is determined by (5.64). Also, (5.64c) determines a hyperplane H in $\mathbb{R}^E$ ((5.64c) is a linear equation with variables in $\mathbb{R}^E$) which is an affine set in $\mathbb{R}^E$ of dimension $|E| - 1$. Then $P_{\text{sptree}}(G) \subseteq H$ and $\dim(P_{\text{sptree}}(G)) + 1 \leq \dim(H) + 1 = |E|$, i.e., $r \leq |E| \leq \binom{|V|}{2}$.

Suppose SwapRound is given a graph $G = (V, E, \psi)$ and a point x in the spanning tree polytope of G. Also, let $\sum_{i=1}^{r} \lambda_i \mathbb{1}_{E(T_i)}$ be the convex combination that SwapRound finds in Line 1 for x. The descriptions of SwapRound and Merge in Algorithm 5.1 and Algorithm 5.2, respectively, stress the manipulation of the spanning trees $T_1, \ldots, T_r$ until they become a single tree T, that is returned. Now we want to analyze another perspective: how the point x evolves in $P_{\text{sptree}}(G)$ until it becomes the incidence vector of a spanning tree of G. Next, by the randomized nature of the algorithm, we will present this progression as a random process, also called a stochastic process, to derive relevant properties about the probability space used by the algorithm.

There are two viewpoints to consider: from SwapRound and Merge. At the level of the SwapRound routine, the point x goes through $r - 1$ changes. First, SwapRound merges spanning trees T_1 and T_2 into a spanning tree $T_{\leq 2}$ and sums their corresponding coefficients so that $x = \sum_{\ell=1}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$ turns into $x = (\lambda_1 + \lambda_2)\mathbb{1}_{E(T_{\leq 2})} + \sum_{\ell=3}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$. At the kth step, for $k \in \{2, \ldots, r - 1\}$, the spanning trees $T_1, \ldots, T_k$ have been merged into a spanning tree $T_{\leq k}$ with corresponding coefficient equal to $\sum_{\ell=1}^{k} \lambda_\ell$; then the spanning trees $T_{\leq k}$ and T_{k+1} are merged into a spanning tree $T_{\leq k+1}$ and their corresponding coefficients are added so that $x = \left(\sum_{\ell=1}^{k} \lambda_\ell\right) \mathbb{1}_{E(T_{\leq k})} + \sum_{\ell=k+1}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$ turns into $x = \left(\sum_{\ell=1}^{k+1} \lambda_\ell\right) \mathbb{1}_{E(T_{\leq k+1})} + \sum_{\ell=k+2}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$. Finally, after $r - 1$ steps like that, we end up with only one spanning tree $T_{\leq r}$ and $x = (\sum_{\ell=1}^{r} \lambda_\ell)\mathbb{1}_{E(T_{\leq r})} = \mathbb{1}_{E(T_{\leq r})}$, i.e., x is the incidence vector of a spanning tree of G.

However, each change of x from SwapRound's point of view is composed of potentially many changes of x from Merge's point of view. More precisely, denote T_1 by $T_{\leq 1}$, and consider when SwapRound merges $T_{\leq k}$ and T_{k+1} into $T_{\leq k+1}$ for an arbitrary $k \in \{1, \ldots, r - 1\}$. To do this, it calls $\text{Merge}(\sum_{\ell=1}^{k} \lambda_\ell, T_{\leq k}, \lambda_{k+1}, T_{k+1})$. Then Merge performs a sequence of so-called **elementary operations**. An elementary operation corresponds to one iteration of the loop in Line 1. Consider $i \in E(T_{\leq k}) \setminus E(T_{k+1})$ and $j \in E(T_{k+1}) \setminus E(T_{\leq k})$ to be the edges swapped. Then this operation on $T_{\leq k}$ and T_{k+1}, with the corresponding coefficients, will modify exactly one of these trees so that they will have one more edge in common (eventually becoming the same spanning tree), and

$$x = \left(\sum_{\ell=1}^{k} \lambda_\ell\right) \mathbb{1}_{E(T_{\leq k})} + \lambda_{k+1} \mathbb{1}_{E(T_{k+1})} + \sum_{\ell=k+2}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$$

will turn into

$$x' = \begin{cases} \left[\left(\sum_{\ell=1}^{k} \lambda_\ell\right)(\mathbb{1}_{E(T_{\leq k})} - e_i + e_j)\right] + \left[\lambda_{k+1} \mathbb{1}_{E(T_{k+1})}\right] + \sum_{\ell=k+2}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)} & \text{if } T_{\leq k} \text{ is modified}\,, \\ \left[\left(\sum_{\ell=1}^{k} \lambda_\ell\right) \mathbb{1}_{E(T_{\leq k})}\right] + \left[\lambda_{k+1}(\mathbb{1}_{E(T_{k+1})} - e_j + e_i)\right] + \sum_{\ell=k+2}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)} & \text{if } T_{k+1} \text{ is modified}\,, \end{cases}$$

i.e., x will have exactly two components in $\mathbb{R}^E$ changed: either components i and j decrease and increase, respectively, by $\sum_{\ell=1}^{k} \lambda_\ell$, or components j and i decrease and increase, respectively, by λ_{k+1}.

Now set $T'_{\leq k}$ and T'_{k+1} to be $T_{\leq k}$ and T_{k+1} after the elementary operation just described, respectively. Note that $T'_{\leq k}$ or T'_{k+1} might be one of the trees $T_{k+2}, \ldots, T_r$. Thus, when we write a convex combination $\sum_{\ell=1}^{m} \mu_\ell \mathbb{1}_{E(T_\ell)}$, the spanning trees $T_1, \ldots, T_m$ are not necessarily pairwise distinct. Independently of that, if $T'_{\leq k} \neq T'_{k+1}$, then another elementary operation will happen between $T'_{\leq k}$ and T'_{k+1} with corresponding coefficients $\sum_{\ell=1}^{k} \lambda_\ell$ and λ_{k+1}. Moreover, this will proceed until the two trees, that start being $T_{\leq k}$ and T_{k+1}, become equal, which happens in $O(|V|)$ iterations (see Theorem 5.16).

Thus, we can describe an arbitrary elementary operation as follows. Given a convex combination $x = \sum_{\ell=1}^{r} \mu_\ell \mathbb{1}_{E(T_\ell)}$ with $T_1, \ldots, T_r$ not necessarily pairwise distinct, an elementary operation between trees T_p, T_q for distinct $p, q \in [r]$ will swap edges $i \in E(T_p) \setminus E(T_q)$ and $j \in E(T_q) \setminus E(T_p)$, and it will turn

$$x = \mu_p \mathbb{1}_{E(T_p)} + \mu_q \mathbb{1}_{E(T_q)} + \sum_{\ell \in [r] \setminus \{p,q\}} \mu_\ell \mathbb{1}_{E(T_\ell)} \tag{5.51}$$

into

$$x' = \begin{cases} \left[\mu_p(\mathbb{1}_{E(T_p)} - e_i + e_j)\right] + \left[\mu_q \mathbb{1}_{E(T_q)}\right] + \sum_{\ell \in [r] \setminus \{p,q\}} \mu_\ell \mathbb{1}_{E(T_\ell)} & \text{if } T_p \text{ is modified}\,, \\ \left[\mu_p \mathbb{1}_{E(T_p)}\right] + \left[\mu_q(\mathbb{1}_{E(T_q)} - e_j + e_i)\right] + \sum_{\ell \in [r] \setminus \{p,q\}} \mu_\ell \mathbb{1}_{E(T_\ell)} & \text{if } T_q \text{ is modified}\,. \end{cases} \tag{5.52}$$

Moreover, note that if we set T'_p and T'_q to be T_p and T_q after the analyzed elementary operation for x, respectively, then we can write vector x' in a more compact form as

$$x' = \mu_p\left(\mathbb{1}_{E(T'_p)} - \mathbb{1}_{E(T_p)}\right) + \mu_q\left(\mathbb{1}_{E(T'_q)} - \mathbb{1}_{E(T_q)}\right) + x\,.$$

Finally, we present the evolution of x as a *stochastic or random process*, that is, a set of random variables on a common probability space that are indexed by some set. Denote by t the total number of elementary operations performed during the algorithm, and set $I \coloneqq \{0, \ldots, t\}$. Then to the sequence of t elementary operations we can associate a random process $\{X_k\}_{k \in I}$ on a probability space $(\Omega, \mathbb{P})$, where for each $k \in I$, we have a multivariate random variable, also called random vector, of the form

$$X_k \colon \omega \in \Omega \mapsto X_k(\omega) \in [0,1]^E\,, \tag{5.53}$$

and for each $e \in E$ we have $X_{k,e} \colon \omega \in \Omega \mapsto [X_k(\omega)]_e$, a random variable on the same probability space. For each $k \in I$, the random vector X_k represents the possible values that the vector x may assume after k elementary operations. Now we want to prove in Theorem 5.20 the correctness of RSR, that is, that the algorithm returns a spanning tree of G and that the last random vector X_t, corresponding to the random spanning tree sampled by SwapRound, satisfies the output conditions in Algorithm 5.1. Actually, the second part we will prove not only for X_t but for each random vector X_k where $k \in I$. We start with the following lemma regarding the expectation of each random vector in this process.

Lemma 5.18. Let $G = (V, E, \psi)$ be a graph, and let $x \in P_{\text{sptree}}(G)$. Suppose SwapRound is given graph G and point x as input. Set t to be the total number of elementary operations that will be performed in the algorithm, and set $I := \{0, \dots, t\}$. Moreover, as in (5.53), define the random process $\{X_k\}_{k \in I}$ on a probability space $(\Omega, \mathbb{P})$ used by the algorithm. Also, suppose X_0 is the constant random variable equal to x. Then we have that $\mathbb{E}[X_k] = x$ for each $k \in \{0, \dots, t\}$. Moreover, for each $k \in \{1, \dots, t\}$, we have that $\mathbb{E}[X_k \mid X_{k-1} = z] = z$ for each $z \in \operatorname{Im}(X_{k-1})$.

Proof. The proof is by induction on k. By hypothesis, $X_0 = x$, and so $\mathbb{E}[X_0] = x$ by item (i) of Proposition 5.6. Suppose $k \geq 1$ and $\mathbb{E}[X_{k-1}] = x$.

Since X_k and X_{k-1} are random vectors on the same probability space, $\mathbb{E}[X_k \mid X_{k-1}]$ is a random vector (see (5.30) for the case involving random variables). Set $X_{k-1}^+ := \{\, z \in \operatorname{Im}(X_{k-1}) : \mathbb{P}(X_{k-1} = z) > 0\}$. Then, from (5.28) of Proposition 5.7,

$$\mathbb{E}[X_k] = \sum_{z \in X_{k-1}^+} \mathbb{E}[X_k \mid X_{k-1} = z]\, \mathbb{P}(X_{k-1} = z)\,.$$

Now consider the k-th elementary operation where we will analyze X_k given that X_{k-1} is determined. Consider a representation of X_{k-1} as a convex combination $X_{k-1} = z := \sum_{\ell=1}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$ for some $\lambda_1, \dots, \lambda_r$ in $\mathbb{R}_+$ with $\sum_{\ell=1}^{r} \lambda_\ell = 1$ and $T_1, \dots, T_r \in \mathcal{T}(G)$. Let $p, q \in [r]$ be integers such that T_p and T_q are the spanning trees involved in this elementary operation. Also, denote by $T_p^{'}$ and $T_q^{'}$ the spanning trees T_p and T_q after this elementary operation, respectively. Note that we are considering an iteration of the loop in Line 3 where $T_1 = T_p$, $T_2 = T_q$, $\beta_1 = \lambda_p$, and $\beta_2 = \lambda_q$. Moreover, let $i \in E(T_p) \setminus E(T_q)$ and let $j \in E(T_q) \setminus E(T_p)$ be the edges to be swapped. Then

$$\begin{aligned}
\mathbb{E}[X_k \mid X_{k-1} = z] &= \mathbb{E}\left[\lambda_p \mathbb{1}_{E(T_p')} + \lambda_q \mathbb{1}_{E(T_q')} + \sum_{\ell \in [r]\setminus\{p,q\}} \lambda_\ell \mathbb{1}_{E(T_\ell)}\right] \\
&= \frac{\lambda_p}{\lambda_p + \lambda_q}\left(\lambda_p \mathbb{1}_{E(T_p)} + \lambda_q(\mathbb{1}_{E(T_q)} - e_j + e_i) + \sum_{\ell \in [r]\setminus\{p,q\}} \lambda_\ell \mathbb{1}_{E(T_\ell)}\right) \\
&\quad + \frac{\lambda_q}{\lambda_p + \lambda_q}\left(\lambda_p(\mathbb{1}_{E(T_p)} - e_i + e_j) + \lambda_q \mathbb{1}_{E(T_q)} + \sum_{\ell \in [r]\setminus\{p,q\}} \lambda_\ell \mathbb{1}_{E(T_\ell)}\right) \\
&= \frac{\lambda_p}{\lambda_p + \lambda_q}(z + \lambda_q(e_i - e_j)) + \frac{\lambda_q}{\lambda_p + \lambda_q}(z - \lambda_p(e_i - e_j)) \qquad (5.54) \\
&= z\,.
\end{aligned}$$

Thus,

$$\mathbb{E}[X_k] = \sum_{z \in X_{k-1}^+} \mathbb{E}[X_k \mid X_{k-1} = z]\, \mathbb{P}(X_{k-1} = z) = \sum_{z \in X_{k-1}^+} z\, \mathbb{P}(X_{k-1} = z) = \mathbb{E}[X_{k-1}] = x$$

where the last equality holds by induction hypothesis. ☐

Lemma 5.18 presents two interesting properties about the probability space used by RSR. Consider the context of the statement of Lemma 5.18. First, the expected value of point x after any $k \in I$ elementary operations of RSR, including the point returned, is x itself. This does not mean the point x always remain unaltered during the algorithm, and then it is returned. It means that if RSR is run multiple times with G and x as input, the "sample average" of the points at each stage of the algorithm (i.e., after any fixed $k \in I$ elementary operations) tends to be x itself.

Second, Lemma 5.18 shows a similar tendency is also true after each elementary operation of the algorithm. That is, given a point z during an intermediate stage of the algorithm, the expected value for z after the next elementary operation is z itself. Moreover, to the algebraic account given in Lemma 5.18 for this fact, we can add a revealing geometric argument. Consider the step (5.54) of proof of Lemma 5.18. There we are analyzing an elementary operation involving a point z and edges i and j. With probability $\lambda_p/(\lambda_p+\lambda_q)$ we change z by λ_q along the direction $e_i - e_j$, while with complementary probability $\lambda_q/(\lambda_p+\lambda_q)$ we change z by λ_p along the opposite direction $e_j - e_i$. However, it turns out that

$$\frac{\lambda_p}{\lambda_p+\lambda_q}\lambda_q(e_i-e_j)+\frac{\lambda_q}{\lambda_p+\lambda_q}\lambda_p(e_j-e_i)=0\,.$$

In other words, the directions of change of z weighted by their respective probabilities cancel each other, and so if we simulate multiple times such elementary operation for z, the "sample average" of the points that z would become would tend to be z itself.

Now the following lemma provides a property that, along with the last lemma, will show the random variables in $\{X_{k,e}\}_{e\in E}$ of random vector X_k are negatively correlated (see (5.40)) for each $k \in I$.

Lemma 5.19. Consider the context of the statement of Lemma 5.18. Let $F \subseteq E$. Then $\mathbb{E}\big[\prod_{e\in F} X_{k,e}\big] \le \prod_{e\in F} x_e$ for each $k \in I$.

Proof. The proof is again by induction on k. Denote $Y_k := \prod_{e\in F} X_{k,e}$ for each $k \in I$.

By hypothesis, $\mathbb{E}[X_0] = x$, and so $\mathbb{E}[Y_0] = \prod_{e\in F} x_e$. Suppose $k \ge 1$ and $\mathbb{E}[Y_{k-1}] \le \prod_{e\in F} x_e$. Also, set $X_{k-1}^{+} := \{\, z \in \mathrm{Im}(X_{k-1}) : \mathbb{P}(X_{k-1}=z) > 0\}$. Then, by the law of total expectation (see (5.28) of Proposition 5.7),

$$\mathbb{E}[Y_k] = \sum_{z\in X_{k-1}^{+}} \mathbb{E}[Y_k \,|\, X_{k-1}=z]\,\mathbb{P}(X_{k-1}=z)\,, \tag{5.55}$$

so we look at $\mathbb{E}[Y_k \,|\, X_{k-1}=z]$. In this k-th elementary operation, the random variable X_{k-1} is a constant $z \in \mathrm{Im}(X_{k-1})$, and denote by i, j the edges from Line 2. Then exactly one of three possibilities must happen: neither i nor j are in F, i.e., $|\{i,j\}\cap F| = 0$; either i or j are in F, i.e., $|\{i,j\}\cap F| = 1$; both i and j are in F, i.e., $|\{i,j\}\cap F| = 2$. Set $S := |\{i,j\}\cap F|$. Note that S is a random variable. Then, by the law of total expectation ((5.29) of Proposition 5.7), for each $z \in X_{k-1}^{+}$ we have

$$\mathbb{E}[Y_k \,|\, X_{k-1}=z] = \sum_{s=0}^{2} \mathbb{E}[Y_k \,|\, S=s\,,\, X_{k-1}=z]\,\mathbb{P}(S=s \,|\, X_{k-1}=z)\,. \tag{5.56}$$

Moreover, note that for each $e \in E$, $s \in \{0,1,2\}$, $z \in \mathrm{Im}(X_{k-1})$, and $z' \in \mathrm{Im}(X_k)$, given the event $X_{k-1} = z$, the events described by the predicates $X_{k,e} = z'_e$ and $S = s$ are conditionally independent (see (5.12)) since F is just an arbitrary chosen subset of E whose knowledge of value S, how many edges of $\{i,j\}$ lie in F, does not alter the probability of $X_{k,e} = z'_e$. Thus, for each $e \in E$ and $s \in \{0,1,2\}$, we have from (5.24) that

$$\begin{aligned}
\mathbb{E}[X_{k,e} \,|\, S=s\,,\, X_{k-1}=z] &= \sum_{z'\in \mathrm{Im}(X_{k,e})} z'_e\,\mathbb{P}(X_{k,e}=z'_e \,|\, S=s\,,\, X_{k-1}=z)\\
&= \sum_{z'\in \mathrm{Im}(X_{k,e})} z'_e\,\mathbb{P}(X_{k,e}=z'_e \,|\, X_{k-1}=z)\\
&= \mathbb{E}[X_{k,e} \,|\, X_{k-1}=z]\,.
\end{aligned} \tag{5.57}$$

Now we analyze the expectation in the RHS of (5.56) for each possible value of S. Suppose $S = 0$. Then $X_{k,e} = X_{k-1,e}$ for each $e \in E$, and so $Y_k = \prod_{e\in F} z_e$ if $X_{k-1} = z \in \mathrm{Im}(X_{k-1})$. Thus,

$$\mathbb{E}[Y_k \,|\, S=0\,, X_{k-1}=z] = \prod_{e\in F} z_e\,. \tag{5.58}$$

Suppose $S = 1$. Also, let f be an edge in $\{i,j\}\cap F$. Then $Y_k = \prod_{e\in F} X_{k,e} = X_{k,f}\prod_{e\in F\setminus\{f\}} X_{k-1,e}$, and so $Y_k = X_{k,f}\prod_{e\in F\setminus\{f\}} z_e$ if $X_{k-1} = z \in \mathrm{Im}(X_{k-1})$. Hence,

$$\begin{aligned}
\mathbb{E}[Y_k \,|\, S=1\,,\, X_{k-1}=z] &= \mathbb{E}[X_{k,f} \,|\, S=1\,,\, X_{k-1}=z]\prod_{e\in F\setminus\{f\}} z_e = \mathbb{E}[X_{k,f} \,|\, X_{k-1}=z]\prod_{e\in F\setminus\{f\}} z_e\\
&= z_f\prod_{e\in F\setminus\{f\}} z_e = \prod_{e\in F} z_e\,,
\end{aligned} \tag{5.59}$$

where the second equality holds by (5.57), and the third equality holds by Lemma 5.18.

Suppose $S = 2$. Then $Y_k = \prod_{e \in F} X_{k,e} = X_{k,i} X_{k,j} \prod_{e \in F \setminus \{i,j\}} X_{k-1,e}$, and so $Y_k = X_{k,i} X_{k,j} \prod_{e \in F \setminus \{i,j\}} z_e$ if $X_{k-1} = z \in \operatorname{Im}(X_{k-1})$. Thus,

$$\mathbb{E}[Y_k \mid S = 2\,,\, X_{k-1} = z] = \mathbb{E}[X_{k,i} X_{k,j} \mid S = 2\,,\, X_{k-1} = z] \prod_{e \in F \setminus \{i,j\}} z_e\,. \tag{5.60}$$

Note that $X_{k,i} X_{k,j} = \frac{1}{4}((X_{k,i} + X_{k,j})^2 - (X_{k,i} - X_{k,j})^2)$. Also, recall the analysis we have made for an arbitrary elementary operation on x as in (5.51) involving edges i and j. There, after the elementary operation, we have a point x' as in (5.52) where either $x'_i + x'_j = (x_i - \mu_p) + (x_j + \mu_p) = x_i + x_j$ or $x'_i + x'_j = (x_i + \mu_q) + (x_j - \mu_q) = x_i + x_j$. Therefore, $X_{k,i} + X_{k,j} = X_{k-1,i} + X_{k-1,j}$, and then

$$\mathbb{E}\big[(X_{k,i} + X_{k,j})^2 \mid S = 2\,,\, X_{k-1} = z\big] = \mathbb{E}\big[(X_{k-1,i} + X_{k-1,j})^2 \mid S = 2\,,\, X_{k-1} = z\big] = (z_i + z_j)^2\,.$$

On the other hand, set $Z := (X_{k,i} - X_{k,j})^2$. From item (v) of Proposition 5.6, if we consider $X = Z$ and $Y = 1$, we have $\mathbb{E}[Z \cdot 1]^2 \le \mathbb{E}[Z^2]\,\mathbb{E}[1]$, i.e., $\mathbb{E}[Z]^2 \le \mathbb{E}[Z^2]$. Thus,

$$\begin{aligned}
\mathbb{E}\big[(X_{k,i} - X_{k,j})^2 \mid S = 2\,,\, X_{k-1} = z\big] &\ge \left(\mathbb{E}[X_{k,i} - X_{k,j} \mid S = 2\,,\, X_{k-1} = z]\right)^2 \\
&= \left(\mathbb{E}[X_{k,i} \mid S = 2\,,\, X_{k-1} = z] - \mathbb{E}[X_{k,j} \mid S = 2\,,\, X_{k-1} = z]\right)^2 \\
&= \left(\mathbb{E}[X_{k,i} \mid X_{k-1} = z] - \mathbb{E}[X_{k,j} \mid X_{k-1} = z]\right)^2 \\
&= (z_i - z_j)^2\,,
\end{aligned}$$

where the second equality holds by (5.57), and the last equality holds by Lemma 5.18. Hence,

$$\begin{aligned}
\mathbb{E}[X_{k,i} X_{k,j} \mid S = 2\,,\, X_{k-1} = z] &= \frac{1}{4}\left(\mathbb{E}\big[(X_{k,i} + X_{k,j})^2 \mid S = 2\,,\, X_{k-1} = z\big]\right. \\
&\qquad \left. - \mathbb{E}\big[(X_{k,i} - X_{k,j})^2 \mid S = 2\,,\, X_{k-1} = z\big]\right) \\
&\le \frac{1}{4}((z_i + z_j)^2 - (z_i - z_j)^2) \\
&= z_i z_j\,,
\end{aligned}$$

and so

$$\mathbb{E}[Y_k \mid S = 2\,,\, X_{k-1} = z] \le \prod_{e \in F} z_e\,. \tag{5.61}$$

From (5.58), (5.59), and (5.61) in (5.56) we have that $\mathbb{E}[Y_k \mid X_{k-1} = z] \le \prod_{e \in F} z_e$. Thus, from (5.55) we have that

$$\mathbb{E}[Y_k] \le \sum_{z \in X^+_{k-1}} \prod_{e \in F} z_e \, \mathbb{P}(X_{k-1} = z) = \sum_{z \in X^+_{k-1}} \mathbb{E}[Y_{k-1} \mid X_{k-1} = z]\, \mathbb{P}(X_{k-1} = z) = \mathbb{E}[Y_{k-1}] \le \prod_{e \in F} x_e\,,$$

where last equality holds by the law of total expectation (see (5.28) of Proposition 5.7), and last inequality holds by induction hypothesis. $\square$

Finally, we use the last two lemmas to prove the correctness of RSR.

Theorem 5.20 (Correctness of Randomized Swap Rounding)**.** *Consider the context of the statement of Lemma 5.18. Then the randomized swap rounding samples a spanning tree of G using the probability space $(\Omega, \mathbb{P})$ that, for each $k \in I$, satisfies*

$$\mathbb{E}[X_k] = x\,, \qquad \text{and} \tag{5.62}$$

$$\mathbb{E}\left[\prod_{e \in S} X_{k,e}\right] \le \prod_{e \in S} x_e \qquad \text{for each } S \subseteq E\,. \tag{5.63}$$

In particular, the random variables in $\{X_{k,e}\}_{e \in E}$ are negatively correlated for each $k \in I$.

Proof. First, we check that the algorithm indeed returns a spanning tree of G. Suppose $x = \sum_{\ell=1}^{r} \lambda_\ell \mathbb{1}_{E(T_\ell)}$ as in Line 1. The algorithm returns an object T that starts being the spanning tree T_1. After this initialization, T is updated $r-1$ times by successive calls of Merge (see Algorithm 5.2). There, after each iteration of loop in Line 3, both T_1 and T_2 are spanning trees of G by construction of edges i and j in Line 2. Moreover, Merge returns T_1. Thus, if Merge is called, it always updates T to become again a spanning tree of G.

Then (5.62) and (5.63) follow directly from Lemma 5.18 and Lemma 5.19, respectively. In addition, from (5.62), we have that $\mathbb{E}[X_{k,e}] = x_e$ for each $k \in I$ and $e \in E$. Thus, with (5.63), it follows that the random variables in $\{X_{k,e}\}_{e\in E}$ are negatively correlated for each $k \in I$. ☐

To finish the section we remark on the importance of the last result for the ApproxATSP (see Algorithm 3.1), namely to perform Line 2. Consider the context of the statement of Lemma 5.18. First, note that the random variables in $\{X_{t,e}\}_{e\in E}$ are 0-1 random variables since RSR always returns an incidence vector of a spanning tree of G. Then, from (5.62), we have $\mathbb{P}(X_{t,e} = 1) = x_e$ for each $e \in E$. Second, we have that the random variables in $\{X_{t,e}\}_{e\in E}$ are negatively correlated. These two properties about the probability space $(\Omega, \mathbb{P})$ are the ones we require in Theorem 5.22 from a randomized algorithm for sampling a spanning tree, and then these properties will help us to prove in Section 5.6 that we can find in polynomial time and using RSR a spanning tree that is (α, s)-thin for desired values of α and s.

5.4 Sampling Random Spanning Tree of G_{z^*}

Let D, x^*, z^*, G_{z^*} be as in Definition 3.2. Recall the goal of this chapter: show we can find a certain spanning tree of G_{z^*}, as in Line 2 of ApproxATSP, in polynomial time. For that we will use the randomized swap rounding (RSR), presented in Section 5.3, that samples spanning trees with a probability space satisfying some desired properties; these properties are detailed in the statement of Theorem 5.22. We will run RSR with G_{z^*} and z^* as input (see Algorithm 5.1); so in Lemma 5.21 we show z^* belongs to the spanning tree polytope of G_{z^*}. We will use the following characterization of the spanning tree polytope of a graph. For a graph $G = (V, E)$, we have that $P_{\text{sptree}}(G)$ is the polytope determined by the following system of inequalities (see [16, Corollary 50.7c.]):

$$z \geq 0\,, \tag{5.64a}$$

$$\mathbb{1}_{E[U]}^{\mathsf{T}} z \leq |U| - 1 \qquad \text{for each } \varnothing \neq U \subsetneq V\,, \tag{5.64b}$$

$$\mathbb{1}_{E}^{\mathsf{T}} z = |V| - 1\,. \tag{5.64c}$$

Lemma 5.21. Let D, x^*, z^*, G_{z^*} be as in Definition 3.2. Then $z^* \in P_{\text{sptree}}(G_{z^*})$.

Proof. Consider the characterization of $P_{\text{sptree}}(G_{z^*})$ given by the system of inequalities in (5.64) for the graph G_{z^*}. Note that by (5.64c), we do not need to consider $U = V$ in (5.64b). By definition of z^* in (3.2) and as $x^* \geq 0$ by (3.1d), we have that z^* satisfies (5.64a).

Now let U be a nonempty and proper subset of V, and set $n := |V|$. On the one hand, $\mathbb{1}_{E[U]}^{\mathsf{T}} z^* \leq \mathbb{1}_{A[U]}^{\mathsf{T}} x^*$ by item (iii) of Proposition 3.3. On the other hand, we have

$$|U| = \sum_{v\in U} \mathbb{1}_{\delta_D^{\text{out}}(v)}^{\mathsf{T}} x^* = \mathbb{1}_{A[U]}^{\mathsf{T}} x^* + \mathbb{1}_{\delta_D^{\text{out}}(U)}^{\mathsf{T}} x^* \geq \mathbb{1}_{A[U]}^{\mathsf{T}} x^* + 1\,,$$

where the first equality holds by (3.1c), and the last inequality holds by (3.1b); i.e., $\mathbb{1}_{A[U]}^{\mathsf{T}} x^* \leq |U| - 1$. Hence, $\mathbb{1}_{E[U]}^{\mathsf{T}} z^* \leq |U| - 1$ and z^* satisfies (5.64b).

Finally, by item (iii) of Proposition 3.3 for $U = V$, we have

$$\mathbb{1}_{E}^{\mathsf{T}} z^* = \left(1 - \frac{1}{n}\right) \mathbb{1}_{A}^{\mathsf{T}} x^* = \left(1 - \frac{1}{n}\right) \sum_{v\in V} \mathbb{1}_{\delta_D^{\text{out}}(v)}^{\mathsf{T}} x^* = \left(1 - \frac{1}{n}\right) n = n - 1\,,$$

where the third equation holds by (3.1c); i.e., z^* satisfies (5.64c). Therefore, z^* satisfies (5.64), and so $z^* \in P_{\text{sptree}}(G_{z^*})$. ☐

In the next result and mainly in Section 5.6, we will be dealing with probability spaces of the form $\mathcal{P} = (\mathcal{T}(G), \mathbb{P})$, where $\mathcal{T}(G)$ is the set of spanning trees of a graph G. Also, it will be of particular interest to compute the probability of events that involve an edge, or a subset of edges, of a graph G being in a spanning tree sampled from $(\mathcal{T}(G), \mathbb{P})$. For instance, in Theorem 5.22, we want to compute the probability of events of the form $\{\, T \in \mathcal{T}(G) : e \in E(T)\}$ for a given edge e of a graph G. Also, in Lemma 5.25, we want to compute the probability of events of the form $\{\, T \in \mathcal{T}(G) : |E(T) \cap \delta(U)| > k\}$ for a graph G, a subset U of $V(G)$, and a $k \in \mathbb{R}_+$. However, as one can see by the statements of these two results, we will not deal directly with such events. We define random variables to help us represent these events (such use of a random variable we have already antecipated when we define random variable and its distribution in Section 5.1).

Let $G = (V, E, \psi)$ be a graph, and let $\mathcal{P} = (\mathcal{T}, \mathbb{P})$ be a probability space where $\mathcal{T} \coloneqq \mathcal{T}(G)$. Define the 0-1 random variables

$$X_e \colon T \in \mathcal{T} \mapsto [e \in E(T)] \qquad \text{for each } e \in E\,, \tag{5.65}$$

with respect to G and $\mathcal{P}$. Then in Theorem 5.22, for instance, we represent the event $\{\, T \in \mathcal{T}(G) : e \in E(T)\}$, for an edge $e \in E$, by the predicate $X_e = 1$ so that $\mathbb{P}(X_e = 1) = \mathbb{P}(\{\, T \in \mathcal{T}(G) : e \in E(T)\})$. Sometimes, one may even encounter the predicate $e \in E(T)$, where T is said to be a *random* spanning tree of G, to indicate this same event.

Theorem 5.22. Let D, c, z^*, G_{z^*} be as in Definition 3.2. Then there is a randomized polynomial-time algorithm that samples a random spanning tree from a probability space $(\mathcal{T}, \mathbb{P})$, where $\mathcal{T} \coloneqq \mathcal{T}(G_{z^*})$, that satisfies what follows. Let $\{X_e\}_{e \in E}$ be random variables defined as in (5.65) with respect to G_{z^*} and $(\mathcal{T}, \mathbb{P})$. Then $\mathbb{P}(X_e = 1) = z^*_e$ for each $e \in E$, and the random variables in $\{X_e\}_{e \in E}$ are negatively correlated.

Proof. Note that for each $e \in E$ we have $\mathbb{P}(X_e = 1) = \mathbb{E}[X_e]$ since X_e is a 0-1 random variable.

By Lemma 5.21, we have z^* in $P_{\text{sptree}}(G_{z^*})$. Then, by Theorem 5.20, $T \coloneqq \text{SwapRound}(G_{z^*}, z^*)$ is a random spanning tree of G_{z^*} such that $\mathbb{E}[X_e] = z^*_e$ for each $e \in E$ and the random variables in $\{X_e\}_{e \in E}$ are negatively correlated. $\square$

5.5 Karger's Bound on the Number of α-Minimum Cuts

In this section, we present a result due to Karger [14]. He developed a simple and elegant algorithm, the **Contraction Algorithm**, to solve the problem of finding a minimum cut in a graph with edge weights. As a consequence, with an algorithm that is a slight modification of the Contraction Algorithm, he found a fact about the combinatorial structure of cuts in a graph, namely an upper bound on the number of cuts within a certain factor of the weight of a minimum cut. We will present the modified algorithm (Algorithm 5.3) and then the bound in Theorem 5.24.

We introduce the following definitions. Let $G = (V, E, \psi)$ be a graph. A function $w \colon E \to \mathbb{R}_+$ is called a **(nonnegative) "weight" function** on the edges of G so that $w_e = w(e)$ is called the *weight* of an edge $e \in E$, and $\mathbb{1}_F^{\mathsf{T}} w$ is called the *weight* of a subset of edges $F \subseteq E$. A subset F of E is a **cut** (of G) if there exists $S \subseteq V$ such that $F = \delta(S)$. If $\varnothing \neq S \subsetneq V$, then $\delta(S)$ is a **nontrivial cut** (of G). We will not consider the cuts $\delta(\varnothing)$ and $\delta(V)$, so from now on, by a cut we mean a nontrivial cut. Also, note that $\varnothing$ is a nontrivial cut if and only if G is disconnected; so, since we will consider only connected graphs, $\varnothing$ will not be a cut. A **minimum cut** of G is a cut of G of minimum weight. A **half-integer** is a number α such that 2α is an integer. We want the following:

> Consider a connected graph G with no loops and $n \coloneqq |V(G)| \geq 2$, a weight function w on the edges of G, and a half-integer $\alpha \geq 1$. Also, denote by $c \geq 0$ the weight of a minimum cut of G. Then we want a polynomial function of n which upper bounds the number of cuts with weight at most $\alpha\, c$. (5.66)

We will create a notation for these special cuts. We say that a cut with weight $\alpha \in \mathbb{R}_+$ times the weight of a minimum cut is an **α-minimum cut**. Moreover, it is worth remarking on the assumptions made for the graph in (5.66); these are the assumptions of Theorem 5.24. Karger showed the bound we present in Theorem 5.24 within the context of the minimum cut problem. In this problem, if the graph is disconnected, then finding a solution is simple: the empty set is always a minimum cut of zero weight, and one can check if the graph is disconnected with a depth-first search, for instance. Thus, Karger supposes the graph is connected. Also, since loops of a graph do not belong to any cut, he ignores them. Finally, he assumes the graph has at least two vertices; otherwise, there will be no cut in the graph.

But why do we care about this bound? Theorem 5.24 enables one to group cuts according to ranges of weight, and then to provide a polynomial upper bound for the number of cuts in each such interval. By "to group cuts according to ranges of weight", we mean, given an interval of nonnegative reals, open or closed, consider all cuts whose weight lies in this interval. We use precisely this idea in Theorem 5.26, one of the theorems that guarantees, with high probability, we can find a desired tree in Line 2 of our main algorithm, Algorithm 3.1.

To find such upper bound, we will use the following idea. Let G be a graph, and denote by $F_1, \dots, F_k$, for some $k \in \mathbb{N}$, the cuts in G whose number we want to bound. Suppose we have a randomized algorithm that samples cuts of G. Let $(\Omega, \mathbb{P})$ be the probability space used by this algorithm, where Ω is the set of cuts of G. Also, suppose this algorithm samples each cut F_i, where $i \in [k]$, with probability at least $p > 0$. Then for the probability of sampling a cut F_i, where $i \in [k]$, we have by (5.3) that

$$\mathbb{P}(\{F_1, \dots, F_k\}) = \sum_{i \in [k]} \mathbb{P}(F_i) \geq \sum_{i \in [k]} p = kp\,.$$

On the other hand,

$$\mathbb{P}(\{F_1, \dots, F_k\}) = \sum_{i \in [k]} \mathbb{P}(F_i) \leq 1\,,$$

where the inequality holds by (5.1) and (5.2). So $kp \leq 1$, whence $k \leq 1/p$. Thus,

> If we have a randomized algorithm that samples each of the cuts, whose number we want to bound, with positive probability, and we have a positive lower bound for this probability, then we will have an upper bound for the number of these cuts. (5.67)

We will apply this idea (5.67) with the following Algorithm 5.3 in the proof of Theorem 5.24. A crucial graph operation employed by both this algorithm and the Contraction Algorithm is *edge contraction*. We start by defining this operation. Let $G = (V, E, \psi)$ be a graph with no loops, and let $w \colon E \to \mathbb{R}_+$ be a weight function. Let $e \in E$ be an edge with $\psi(e) = uv$ for some $u, v \in V$. Contracting the edge e means creating a new graph, denoted by G/e, that is equal to G except for three changes: the edges between u and v are removed; the vertices u, v are replaced by a new vertex z, not in V; the set of edges incident to either u or v becomes the set of edges incident to z. Thus, note that after the contraction of e the resulting graph can have parallel edges but no loops. More precisely, the graph G/e that we denote by $G' = (V', E', \psi')$ is defined by

$$V' \coloneqq V \setminus \{u, v\} \cup \{z\} \quad \text{with } z \notin V\,, \tag{5.68}$$

$$E' \coloneqq E[V \setminus \{u, v\}] \cup \delta_G(\{u, v\})\,, \tag{5.69}$$

and

$$\psi'(f) \coloneqq \begin{cases} \psi(f) & \text{for each } f \in E[V \setminus \{u, v\}]\,, \\ yz & \text{for each } f \in \delta_G(\{u, v\}) \text{ s.t. } \psi(f) = xy \text{ with } x \in \{u, v\} \text{ and } y \in V \setminus \{u, v\}\,. \end{cases} \tag{5.70}$$

Note that the edge set of G/e is a subset of the edge set of G. Hence, we can compute the weight of any edge or subset of edges of G/e using the weight function w.

Now we describe the algorithm. Given a connected graph G with no loops and at least two vertices, a weight function w on the edges of G, and a half-integer $\alpha \geq 1$, the algorithm is divided into two parts. First, it randomly contracts edges of G until G has exactly 2α vertices (actually, note that $|V(G)|$ can be less than 2α; this special case we treat in the proof of Theorem 5.24, and it will have the same bound for the case we are analyzing here). Denote the resulting graph after these contractions by G'. Then the algorithm samples uniformly at random a nonempty and proper subset S of vertices of $V(G')$. That is, since G' has 2α vertices, it has $2^{2\alpha} - 2$ nonempty and proper subsets of vertices, and so each such subset is selected with probability $1/(2^{2\alpha} - 2)$. Finally, it outputs the set $\delta(S)$. The compact description is as follows.

Algorithm 5.3: SampleαMinimumCuts(G, w, α)

Input:
(i) a connected graph $G = (V, E, \psi)$ with no loops and at least two vertices,
(ii) a weight function $w\colon E \to \mathbb{R}_+$, and
(iii) a half-integer $\alpha \geq 1$.

Output: A cut of G.

1. **while** G has more than 2α vertices **do**
2. Sample an edge e with probability proportional to the weight of e
3. $G \leftarrow G/e$
4. Sample uniformly at random a proper and nonempty subset S of $V(G)$
5. **return** $\delta(S)$

The next lemma reveals three properties about edge contraction that are crucial for the correctness of Algorithm 5.3 and for the proof of Theorem 5.24. item (i) is used to prove the other two results. items (ii) and (iii) will be used in the proof of Theorem 5.24. item (ii) gives a necessary and sufficient condition for a cut to "survive" an edge contraction, from which we derive the condition a cut must satisfy in the first part of the algorithm, the part with edge contractions, to have a chance of being sampled in the second part. item (iii) justifies a lower bound for the weight of a minimum cut in any graph of the first part of the algorithm, which ultimately helps us find a lower bound for the probability that a cut "survives" the first part of the algorithm. Moreover, item (iii) guarantees that the output of Algorithm 5.3 with input G is always a cut of G.

Lemma 5.23 (Cut Preservation)**.** Let $G = (V, E, \psi)$ be a graph with no loops, and let $e \in E$ be an edge with $\psi(e) = uv$ for some $u, v \in V$. Denote by $G' = (V', E', \psi')$ the graph G/e, and denote by z the vertex, not in V, that replaces u, v after the contraction of e in G (see (5.68)). Then we have the following:

(i) Let S be any subset of V that contains u and v. Also, set $S' := (S \setminus \{u, v\}) \cup \{z\}$, i.e., $S = (S' \setminus \{z\}) \cup \{u, v\}$. Then $\delta_G(S) = \delta_{G'}(S')$. Moreover, we have $\delta_G\big(\overline{S}\big) = \delta_{G'}\big(\overline{S'}\big)$.

(Any subset S of vertices of G that either has both ends of the edge to be contracted or none will have its cut $\delta_G(S)$ preserved after the edge contraction.)

(ii) Let $F \subseteq E$ be a cut of G. Then F is a cut of G/e if and only if $e \notin F$.

(Necessary and sufficient condition for a cut to "survive" an edge contraction.)

(iii) Let $F \subseteq E'$ be a cut of G'. Then F is a cut of G.

(Every cut of the contracted graph is a cut of the original graph.)

Proof. **(i)** Let $f \in E$. Then

$$\begin{aligned}
f \in \delta_G(S) &\Leftrightarrow \psi(f) = xy \text{ s.t. } x \in S \text{ and } y \in V \setminus S \\
&\Leftrightarrow (\psi(f) = xy \text{ s.t. } x \in \{u,v\} \text{ and } y \in V \setminus S) \text{ or } (\psi(f) = xy \text{ s.t. } x \in S \setminus \{u,v\} \text{ and } y \in V \setminus S) \\
&\Leftrightarrow (\psi'(f) = zy \text{ s.t. } y \in V \setminus S) \qquad\qquad\quad \text{or } (\psi'(f) = xy \text{ s.t. } x \in S \setminus \{u,v\} \text{ and } y \in V \setminus S) \\
&\Leftrightarrow (\psi'(f) = zy \text{ s.t. } y \in V' \setminus S') \qquad\qquad\quad \text{or } (\psi'(f) = xy \text{ s.t. } x \in S' \setminus \{z\} \text{ and } y \in V' \setminus S') \\
&\Leftrightarrow \psi'(f) = xy \text{ s.t. } x \in S' \text{ and } y \in V' \setminus S' \\
&\Leftrightarrow f \in \delta_{G'}(S')\,,
\end{aligned}$$

where the third equivalence holds by definition of ψ' (see (5.70)), and the fourth equivalence holds by definition of V, V', S, and S'. Moreover, since $\delta_G(S) = \delta_G\big(\overline{S}\big)$ and $\delta_{G'}(S') = \delta_{G'}\big(\overline{S'}\big)$, we have $\delta_G\big(\overline{S}\big) = \delta_{G'}\big(\overline{S'}\big)$.

(ii) ($\Rightarrow$) Since $e \notin E'$ and $F \subseteq E'$, we have $e \notin F$.

($\Leftarrow$) Since F is a cut of G, there exists a subset S of V such that $\delta_G(S) = F$. Since $e \notin F$, we have either $e \in E[S]$ or $E[\overline{S}\,]$. Suppose, without loss of generality, $e \in E[S]$, and set $S' := (S \setminus \{u, v\}) \cup \{z\}$. Then, by item (i), $\delta_G(S) = \delta_{G'}(S')$, i.e., F is a cut of G'.

(iii) Since F is a cut of G', there exists a subset S' of V' such that $\delta_{G'}(S') = F$. Then either $z \in S'$ or $z \in \overline{S'}$. Suppose, without loss of generality, $z \in S'$, and set $S := (S' \setminus \{z\}) \cup \{u, v\}$. Then, by item (i), $\delta_{G'}(S') = \delta_G(S)$, i.e., F is a cut of G. $\square$

Theorem 5.24 (Bound on the number of α-minimum cuts, [14, Theorem 6.2])**.** Let $G = (V, E, \psi)$ be a connected graph with no loops and $n := |V| \geq 2$, and let $w\colon E \to \mathbb{R}_+$ be a weight function. Let $\alpha \geq 1$ be a half-integer, i.e., $2\alpha \in \mathbb{Z}$. Also, set c to be the weight of a minimum cut. Then the number of α-minimum cuts in G is at most $n^{2\alpha}$. Moreover, if $\alpha < n/2$ this number is at most $2^{2\alpha-1}\binom{n}{2\alpha}$, which is less than $n^{2\alpha}$.

Proof. We divide into two cases. First, suppose $\alpha \geq n/2$. Then the desired bound follows immediately from a bound on the total number of cuts in G. Indeed, there exist $2^n - 2$ nonempty and proper subsets of V, and each cut is determined by at least two subsets of vertices. So there exist at most $2^{n-1} - 1$ cuts in G (actually, since G is connected, one can show each cut is determined by exactly two subsets of vertices, and so this bound is exact). Also, $n \geq 2$ and $n - 1 < n \leq 2\alpha$. Hence, $2^{n-1} - 1 < n^{2\alpha}$.

Now suppose $\alpha < n/2$. We apply the idea (5.67) with the algorithm Algorithm 5.3. So first, we show

$$\text{Algorithm 5.3 samples each } \alpha\text{-minimum cut with positive probability.} \tag{5.71}$$

Recall that Algorithm 5.3 has two parts: first, the algorithm has a sequence of edge contractions until the graph has 2α vertices; then the algorithm makes a uniformly random sampling of a cut of the resulting graph after the edge contractions. Then, as long as a cut "survives" the first part, it has a positive probability of being sampled. We show that the probability of an α-minimum cut "surviving" the first part is positive.

Let $F \subseteq E$ be a cut of G with weight at most $\alpha\, c$. By item (ii) of Lemma 5.23, a cut "survives" an edge contraction of an edge e if and only if $e \notin F$. So F will "survive" the first part of the algorithm if it "survives" after each edge contraction that happens. Since each edge contraction decreases by one the number of vertices, the input graph has n vertices, and after the edge contractions there are exactly 2α vertices, there will be $n - 2\alpha$ edge contractions. So denote the edges to be contracted by $e_1, \dots, e_{n-2\alpha}$. We want to determine $\mathbb{P}\big(\wedge_{i\in[n-2\alpha]}(e_i \notin F)\big)$. By Proposition 5.2, we have

$$\mathbb{P}\big(\wedge_{i\in[n-2\alpha]}(e_i \notin F)\big) = \mathbb{P}(e_1 \notin F) \cdot \mathbb{P}(e_2 \notin F \,|\, e_1 \notin F) \cdots \mathbb{P}\big(e_{n-2\alpha} \notin F \,|\, \wedge_{i\in[n-2\alpha-1]}\, (e_i \notin F)\big)\,.$$

Now let us analyze an arbitrary edge contraction in the algorithm. Consider an arbitrary iteration in the loop of Line 1 of Algorithm 5.3 with input G, and denote the current graph by G'. Also, suppose G' has r vertices, where $2\alpha < r \leq n$, and that the cut F is intact, i.e., no edge of F was contracted so far. By item (iii) of Lemma 5.23, every cut of G' is a cut of G, and then the weight of a minimum cut of G' is at least c. Hence, the total weight of G' is

$$\frac{1}{2} \sum_{v\in V(G')} \mathbb{1}^{\mathsf{T}}_{\delta_{G'}(v)} w \geq \frac{1}{2} \sum_{v\in V(G')} c = \frac{cr}{2}\,.$$

On the other hand, the weight of F is at most αc. Then the probability of choosing an edge of F to contract in this iteration is at most $(\alpha c)/(rc/2) = 2\alpha/r$, and so the probability of not choosing an edge of F is at least $1 - 2\alpha/r$. Thus,

$$\begin{aligned}
\mathbb{P}\big(\wedge_{i\in[n-2\alpha]}(e_i \notin F)\big) &\geq \left(1 - \frac{2\alpha}{n}\right)\left(1 - \frac{2\alpha}{n-1}\right)\cdots\left(1 - \frac{2\alpha}{2\alpha+1}\right) \\
&= \left(\frac{n-2\alpha}{n}\right)\left(\frac{n-2\alpha-1}{n-1}\right)\cdots\left(\frac{1}{2\alpha+1}\right) \\
&= \frac{(n-2\alpha)!(2\alpha)!}{n!} = \binom{n}{2\alpha}^{-1},
\end{aligned}$$

which is a positive real, and so we have (5.71).

Now we complete the analysis and give a positive lower bound for the probability of F being sampled by the algorithm. Given that the cut F "survives" the first part, there exist at least two subsets that determine

F out of a total of $2^{2\alpha}-2$ proper and nonempty subsets of the graph resulting from the first part (actually, one can show that there are exactly two since the graph is connected). Denote by S the subset sampled in the second part. Then the probability that the cut F is returned by the algorithm is

$$\mathbb{P}(F) = \mathbb{P}\big(\wedge_{i\in[n-2\alpha]}(e_i \notin F)\big)\,\mathbb{P}(\delta(S)=F) \geq \binom{n}{2\alpha}^{-1}\frac{2}{2^{2\alpha}-2} = \binom{n}{2\alpha}^{-1}\frac{1}{2^{2\alpha-1}} = \frac{(n-2\alpha)!(2\alpha)!}{n!}\cdot\frac{1}{2^{2\alpha-1}}$$
$$= \frac{(2\alpha)\,(2\alpha-1)\cdots 2\cdot 1}{n\,(n-1)\,(n-2)\cdots(n-2\alpha-1)}\cdot\frac{1}{2^{2\alpha-1}} = \frac{1}{n}\cdot\frac{2\alpha}{2(n-1)}\cdot\frac{2\alpha-1}{2(n-2)}\cdots\frac{2}{2(n-2\alpha+1)} \geq \frac{1}{n^{2\alpha}} =: p\,,$$

where the last inequality holds since 2α is an integer greater than or equal to 2, and so we have in the LHS a product of 2α factors, each one greater than or equal to $1/n$.

In other words, an arbitrary cut, of those whose number we want to bound, is sampled using Algorithm 5.3 with probability at least $p > 0$. Finally, we complete the use of idea (5.67). The number of α-minimum cuts is at most $1/p = n^{2\alpha}$. □

5.6 Finding an $(\alpha, 2)$-Thin Tree of G_{z^*} With High Probability

Let D, c, z^*, G_{z^*} be as in Definition 3.2, and let α be a positive real. Theorem 5.22 presents a randomized polynomial-time algorithm $\mathcal{A}$ that samples spanning trees of G_{z^*} using a probability space with certain properties. We show that, by using such an algorithm, one can find in polynomial time a spanning tree of G_{z^*} that is $(\alpha, 2)$-thin (see Definition 3.2) with high probability.

From Definition 3.2, the spanning tree we search for must satisfy two properties: it must be α-thin with respect to z^* with high probability; it must have cost, with cost function c^* as in (3.3), of at most $2\,\mathrm{OPT}_{\mathrm{HK}}$ with high probability. The first requirement is satisfied by directly using $\mathcal{A}$ since Theorem 5.26 shows that any spanning tree of G_{z^*} sampled by $\mathcal{A}$ is α-thin with respect to z^* with high probability, namely at least $1-1/(n-1)$. The second requirement demands a bit more. First, Theorem 5.27 shows that any spanning tree of G_{z^*} sampled using $\mathcal{A}$ has cost at most $2\,\mathrm{OPT}_{\mathrm{HK}}$ with probability at least $1/2$. Then Corollary 5.28 shows that if we sample $\lceil 2\ln n\rceil$ spanning trees using $\mathcal{A}$, and we pick one with minimum cost, this chosen tree will have the desired cost with high probability, namely at least $1-1/n > 1-1/(n-1)$. Finally, Theorem 5.29 joins Theorem 5.26 and Corollary 5.28 to show that indeed we can find a desired spanning tree of G_{z^*} with high probability, namely greater than $1-2/(n-1)$.

We start with the first requirement. Denote by T the random spanning tree of G_{z^*} we will sample. Roughly, for T to be α-thin with respect to z^*, we need no cut of G_{z^*} to contain too many edges of T. But how many edges is too much? Well, by Definition 3.1, for any cut, T can have at most as many edges as α times the sum of the entries of z^* that are edges of the cut. To prove T satisfies this requirement with high probability, we first show, in Lemma 5.25, a bound for the probability that T violates the α-thiness property for an arbitrarily chosen cut. Then, in Theorem 5.26, we bound the probability that there exists a cut of G_{z^*} for which T violates the α-thiness property.

Finally, we remark on the events we will analyze. Since T is a random spanning tree, to describe events that involve its edges we will use the random variables $\{X_e\}_{e\in E}$ defined as in (5.65) with respect to G_{z^*} and a probability space to be introduced. Thus, in an event description involving $|E(T)\cap\delta(U)|$ (from Definition 3.1), for a subset U of V, such term will be replaced by the sum $\sum_{e\in\delta(U)} X_e$. Moreover, we take advantage of the family notation, described in [13, Section 9], for instance, where $\{X_e\}_{e\in E}$ is the function $X\colon E\to X_e$, and so we write $\sum_{e\in\delta(U)} X_e = \mathbb{1}_{\delta(U)}^{\mathsf{T}} X$.

Lemma 5.25. Let D, c, z^*, G_{z^*} be as in Definition 3.2. Suppose G_{z^*} has $n\geq 3$ vertices. Let $(\mathcal{T},\mathbb{P})$ be the probability space as in the statement of Theorem 5.22 from which one can sample spanning trees of G_{z^*}. Let $\{X_e\}_{e\in E}$ be random variables defined as in (5.65) with respect to G_{z^*} and $(\mathcal{T},\mathbb{P})$. Then, for each U a nonempty and proper subset of V, one has

$$\mathbb{P}\Big(\mathbb{1}_{\delta(U)}^{\mathsf{T}} X > \alpha\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*\Big) \leq n^{-2.5\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*}\,, \tag{5.72}$$

where $\alpha := 4\ln n/\ln\ln n$.

Proof. Let U be a nonempty and proper subset of V. Since the random variables in $\{X_e\}_{e\in E}$ are negatively correlated by Theorem 5.22, we show there exists a scalar $\beta > 0$ such that

$$\mathbb{P}\Big(\mathbb{1}_{\delta(U)}^{\mathsf{T}} X > \alpha \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*\Big) = \mathbb{P}\Big(\mathbb{1}_{\delta(U)}^{\mathsf{T}} X > (1+\beta)\mathbb{E}\Big[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X\Big]\Big) \tag{5.73}$$

so that we can apply the Chernoff bound of Theorem 5.14 to the RHS of (5.73). Let $e \in E$. Since X_e is a 0-1 random variable, we have $\mathbb{E}(X_e) = \mathbb{P}(X_e)$. Also, $\mathbb{P}(X_e = 1) = z_e^*$ by Theorem 5.22. Thus, $\mathbb{E}[X_e] = z_e^*$, and so, by linearity of expectation (see item (ii) of Proposition 5.6),

$$\mathbb{E}\Big[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X\Big] = \sum_{f\in\delta(U)} \mathbb{E}[X_f] = \sum_{f\in\delta(U)} z_f^* = \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^* \,. \tag{5.74}$$

As $\alpha > 1$ for $n \geq 3$, and by (5.74), set $\beta \coloneqq \alpha - 1 > 0$ so that $(1+\beta)\mathbb{E}\Big[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X\Big] = \alpha \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*$, and then (5.73) follows. Now, as we said, we apply Theorem 5.14 to the RHS of (5.73), and we obtain

$$\begin{aligned}
\mathbb{P}\Big(\mathbb{1}_{\delta(U)}^{\mathsf{T}} X > \alpha \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*\Big) &\leq \left(\frac{e^{\beta}}{(1+\beta)^{1+\beta}}\right)^{\mathbb{E}[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X]} \leq \left(\frac{e^{1+\beta}}{(1+\beta)^{1+\beta}}\right)^{\mathbb{E}[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X]} \\
&= \left(\frac{e}{1+\beta}\right)^{(1+\beta)\mathbb{E}[\mathbb{1}_{\delta(U)}^{\mathsf{T}} X]} = \left(\frac{e}{1+\beta}\right)^{\alpha \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*} \\
&= \left[\left(\frac{e}{\alpha}\right)^{\alpha}\right]^{\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*} .
\end{aligned} \tag{5.75}$$

Since $\alpha = 4\ln n/\ln\ln n$, we have

$$\begin{aligned}
\ln\Big[\Big(\frac{e}{\alpha}\Big)^{\alpha}\Big] &= \frac{4\ln n}{\ln\ln n}\ln\Big(e\frac{\ln\ln n}{4\ln n}\Big) \\
&= \frac{4\ln n}{\ln\ln n}[\ln e/4 - \ln\ln n + \ln\ln\ln n] \\
&= 4\ln n\left[\frac{\ln e/4}{\ln\ln n} - 1 + \frac{\ln\ln\ln n}{\ln\ln n}\right] \\
&\leq -4\ln n\Big(1 - \frac{\ln\ln\ln n}{\ln\ln n}\Big) && \text{since } e < 4 \text{ and } \ln\ln n > 0 \text{ for } n \geq 3 \\
&\leq -4\Big(1-\frac{1}{e}\Big)\ln n && \text{since } \frac{\ln\ln\ln n}{\ln\ln n} \leq \frac{1}{e} \text{ for each } n \geq 3 \\
&\leq -2.5\ln n \\
&= \ln n^{-2.5} ,
\end{aligned} \tag{5.76}$$

As for each $x, y \in \mathbb{R}_{++}$ we have that $\ln x \leq \ln y$ implies $x \leq y$, it follows from (5.76) and $\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^* > 0$ that $[(e/\alpha)^{\alpha}]^{\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*} \leq n^{-2.5\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*}$. Then, by (5.75),

$$\mathbb{P}\Big(\mathbb{1}_{\delta(U)}^{\mathsf{T}} X > \alpha \mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*\Big) \leq n^{-2.5\mathbb{1}_{\delta(U)}^{\mathsf{T}} z^*} . \qquad \square$$

Theorem 5.26. Let D, c, x^*, z^*, G_{z^*} be as in Definition 3.2. Suppose G_{z^*} has $n \geq 5$ vertices. Let $(\mathcal{T}, \mathbb{P})$ be the probability space as in the statement of Theorem 5.22 from which one can sample spanning trees of G_{z^*}. Let $\{X_e\}_{e\in E}$ be random variables defined as in (5.65) with respect to G_{z^*} and $(\mathcal{T}, \mathbb{P})$. Moreover, denote by $\mathcal{C}$ the set of nontrivial cuts of G_{z^*}. Then

$$\mathbb{P}\big(\mathbb{1}_F^{\mathsf{T}} X > \alpha\, \mathbb{1}_F^{\mathsf{T}} z^* \text{ for some } F \in \mathcal{C}\big) \leq \frac{1}{n-1} ,$$

where $\alpha \coloneqq 4\ln n/\ln\ln n$.

Proof. By the union bound (see item (ii) from Proposition 5.1), it follows that

$$\mathbb{P}\big(\mathbb{1}_F^\intercal X > \alpha \mathbb{1}_F^\intercal z^* \text{ for some } F \in \mathcal{C}\big) \le \sum_{F\in\mathcal{C}} \mathbb{P}\big(\mathbb{1}_F^\intercal X > \alpha \mathbb{1}_F^\intercal z^*\big)\,. \tag{5.77}$$

The graph G_{z^*} is connected by Proposition 3.4. Also, G_{z^*} has no loops since D is a complete digraph (see definitions in Section 2.3), and G_{z^*} has more than two vertices. Thus, G_{z^*} satisfies the hypotheses of Theorem 5.24, and then instead of applying directly Lemma 5.25 to the RHS of the last inequality, we use Theorem 5.24. This result shows that the number of cuts in G_{z^*} with weight at most β times the weight of a minimum cut of G_{z^*} is at most $n^{2\beta}$ for any half-integer $\beta \ge 1$. First, we find the weight of a minimum cut of G_{z^*} with weight function z^*. By item (ii) of Proposition 3.3, for any nonempty and proper subset U of V we have

$$\mathbb{1}_{\delta(U)}^\intercal z^* = 2\left(1-\frac{1}{n}\right)\mathbb{1}_{\delta_D^{\text{out}}(U)}^\intercal x^* \ge 2\left(1-\frac{1}{n}\right), \tag{5.78}$$

where the last inequality holds by (3.1b). If U is a singleton, then $\mathbb{1}_{\delta_D^{\text{out}}(U)}^\intercal x^* = 1$ by (3.1c), and hence the inequality in (5.78) is an equality. Then $2(1-1/n)$ is the weight of a minimum cut of G_{z^*}.

Set $\mathcal{C}_i := \{\, F \in \mathcal{C} : \mathbb{1}_F^\intercal z^* \in [(i-1)(1-1/n), i(1-1/n)]\}$ for each $i \ge 3$. Note that for each $F \in \mathcal{C}$ there is at least one $i \ge 3$ such that $F \in \mathcal{C}_i$. Moreover, since for each integer $i \ge 2$ there are n^i cuts in G_{z^*} with weight at most $\frac{i}{2}(2(1-1/n)) = i(1-1/n)$ by Theorem 5.24, we have that $|\mathcal{C}_i| \le n^i$ for each integer $i \ge 3$. Then

$$\begin{aligned}
\sum_{F\in\mathcal{C}} \mathbb{P}\big(\mathbb{1}_F^\intercal X > \alpha \mathbb{1}_F^\intercal z^*\big) &\le \sum_{i=3}^{\infty}\sum_{F\in\mathcal{C}_i} \mathbb{P}\big(\mathbb{1}_F^\intercal X > \alpha \mathbb{1}_F^\intercal z^*\big) \le \sum_{i=3}^{\infty}\sum_{F\in\mathcal{C}_i} n^{-2.5\mathbb{1}_F^\intercal z^*} && \text{by Lemma 5.25}\\
&\le \sum_{i=3}^{\infty}\sum_{F\in\mathcal{C}_i} n^{-2.5(i-1)(1-1/n)} \quad = \sum_{i=3}^{\infty} |\mathcal{C}_i| n^{-2.5(i-1)(1-1/n)}\\
&\le \sum_{i=3}^{\infty} n^i n^{-2.5(i-1)(1-1/n)} \quad = \sum_{i=3}^{\infty} n^{(-1.5i+(2.5i)/n)+(2.5-2.5/n)}\\
&\le \sum_{i=3}^{\infty} n^{-i+2} \quad = \sum_{i=1}^{\infty} \frac{1}{n^i} =: S\,,
\end{aligned} \tag{5.79}$$

where the last inequality holds since $-1.5i + (2.5i)/n \le -i$ and $2.5 - 2.5/n \le 2$ for $n \ge 5$. Note that S is a geometric series, a series with the same constant ratio between successive terms; the ratio for the series S is $1/n$. Then

$$S - \frac{1}{n}S = \frac{1}{n} \Rightarrow \left(1-\frac{1}{n}\right)S = \frac{1}{n} \Rightarrow S = \frac{1}{n-1}\,. \qquad \square$$

Consider the context of the statement of Theorem 5.26. With the last two results we have shown that a tree sampled by the algorithm provided by Theorem 5.22 satisfies the first requirement we have established at the beginning of this section, that is, it is α-thin with respect to z^* with high probability, namely at least $1 - 1/(n-1)$.

Before proceeding with the other results for the second requirement, we want to "justify", or at least provide an intuition for, using Karger's result (Theorem 5.24) in the proof of the last theorem. Suppose that we had directly applied Lemma 5.25 to the RHS of (5.77); then we would have had

$$\sum_{F\in\mathcal{C}} \mathbb{P}\big(\mathbb{1}_F^\intercal X > \alpha \mathbb{1}_F^\intercal z^*\big) = \sum_{F\in\mathcal{C}} n^{-2.5\mathbb{1}_F^\intercal z^*}\,.$$

From the proof of Theorem 5.26, recall that the weight of a minimum cut of G_{z^*} is $2(1-1/n)$, i.e., among all non-trivial cuts of G_{z^*} at least one has the weight $2(1-1/n)$, and all others have at least this weight. Moreover, note that there are $2^n - 2$ nonempty and proper subsets of V, and one can show that exactly each two of them determine a cut of G_{z^*} since this graph is connected; as a consequence, there are $2^{n-1} - 1$ cuts in G_{z^*}. Then a reasonable, or one could say natural, way to develop the last equation would be

$$\sum_{F\in\mathcal{C}} n^{-2.5\mathbb{1}_F^\intercal z^*} \le \sum_{F\in\mathcal{C}} n^{-5(1-1/n)} = \frac{2^{n-1}-1}{n^{5(1-1/n)}} \ge \frac{2^{n-2}}{n^5}$$

where the last inequality holds as $n \geq 2$. So following a very plausible reasoning results in a fraction with an inconvenient exponential of n in the numerator. Now consider the sequence of equations and inequalities that end in (5.79). By organizing the cuts in weight ranges with respect to the weight of a minimum cut, we obtain a dreadful series after the first inequality. However, after the third inequality, as we have gained the information of the weight range each cut belongs to, we can assign a higher lower bound for the weight of potentially many cuts than the weight of a minimum cut. Moreover, after the fourth inequality, by Karger's result, we have a polynomial function of n bound for the number of cuts in each weight range instead of an exponential of n. As the proof shows, these last two facts combined are enough to produce a geometric series whose terms are less than 1 and that converges to $1/(n-1)$, a real smaller than 1 that decreases as n increases.

Now with the next two results we show how to sample, in polynomial time and using the algorithm provided by Theorem 5.22, a tree that satisfies the second established requirement: it must have cost, with respect to cost function c^* as in (3.3), of at most $2\,\mathrm{OPT}_{\mathrm{HK}}$ with high probability.

Theorem 5.27. Let D, c, x^*, z^*, G_{z^*}, c^*, $\mathrm{OPT}_{\mathrm{HK}}$ be as in Definition 3.2. Suppose G_{z^*} has $n \geq 5$ vertices. Let $(\mathcal{T}, \mathbb{P})$ be the probability space as in the statement of Theorem 5.22 from which one can sample spanning trees of G_{z^*}. Let $\{X_e\}_{e \in E}$ be random variables defined as in (5.65) with respect to G_{z^*} and $(\mathcal{T}, \mathbb{P})$. Moreover, let $C\colon \mathcal{T} \to \mathbb{R}$ be a random variable such that $C(T) \coloneqq \mathbb{1}_{E(T)}^{\mathsf{T}} c^*$ for each $T \in \mathcal{T}$. Then

$$\mathbb{P}(C > 2\,\mathrm{OPT}_{\mathrm{HK}}) < \frac{1}{2}\,.$$

Proof. By the equivalent expression for expectation (see (5.15)),

$$\begin{aligned}\mathbb{E}[C] &= \sum_{T \in \mathcal{T}} C(T)\,\mathbb{P}(T) = \sum_{T \in \mathcal{T}} \sum_{e \in E} [e \in E(T)]\, c^*_e\, \mathbb{P}(T) = \sum_{e \in E} c^*_e \sum_{T \in \mathcal{T}} [e \in E(T)]\, \mathbb{P}(T) \\ &\overset{(5.3)}{=} \sum_{e \in E} c^*_e\, \mathbb{P}(\{\, T \in \mathcal{T} : e \in E(T) \}) = \sum_{e \in E} c^*_e\, \mathbb{P}(X_e = 1)\,.\end{aligned}$$

By Theorem 5.22, $\mathbb{P}(X_e = 1) = z^*_e$ for each $e \in E$, so it follows

$$\begin{aligned}\mathbb{E}[C] &= \sum_{\{u,v\} \in E} c^*_{\{u,v\}}\, z^*_{\{u,v\}} \overset{(3.2)}{=} \sum_{\{u,v\} \in E} c^*_{\{u,v\}}\, \frac{n-1}{n} (x^*_{uv} + x^*_{vu}) \overset{(3.3)}{\leq} \frac{n-1}{n} \sum_{\{u,v\} \in E} (c_{uv}\, x^*_{uv} + c_{vu}\, x^*_{vu}) \\ &= \frac{n-1}{n} \sum_{u,v \in V} (c_{uv}\, x^*_{uv} + c_{vu}\, x^*_{vu}) = \frac{n-1}{n} \sum_{a \in A} c_a\, x^*_a < \mathrm{OPT}_{\mathrm{HK}}\,,\end{aligned}$$

where the third equality holds as D is a complete digraph and by item (i) of Proposition 3.3, and the last inequality holds since x^* is an optimum solution of the Held-Karp relaxation of (mATSP) determined by D and c. Then, by Markov's inequality (see Theorem 5.8),

$$\mathbb{P}(C > 2\,\mathrm{OPT}_{\mathrm{HK}}) \leq \frac{\mathbb{E}[C]}{2\,\mathrm{OPT}_{\mathrm{HK}}} < \frac{\mathrm{OPT}_{\mathrm{HK}}}{2\,\mathrm{OPT}_{\mathrm{HK}}} = \frac{1}{2}\,. \qquad \square$$

Corollary 5.28. Let D, c, z^*, G_{z^*} be as in Definition 3.2. Suppose G_{z^*} has $n \geq 5$ vertices. Let $(\mathcal{T}, \mathbb{P})$ be the probability space as in the statement of Theorem 5.22 from which one can sample spanning trees of G_{z^*}. Let $\{X_e\}_{e \in E}$ be random variables defined as in (5.65) with respect to G_{z^*} and $(\mathcal{T}, \mathbb{P})$. Set $k \coloneqq \lceil 2 \ln n \rceil$, and let $T_1, \ldots, T_k$ be the result of k independent samplings from $(\mathcal{T}, \mathbb{P})$. Moreover, let T^* be a tree among the k sampled that has minimum cost. Then

$$\mathbb{P}\Big(\mathbb{1}_{E(T^*)}^{\mathsf{T}} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}}\Big) < \frac{1}{n}\,.$$

Proof. Since $\mathbb{1}_{E(T^*)}^\mathsf{T} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}}$ if and only if $\mathbb{1}_{E(T_i)}^\mathsf{T} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}}$ for each $i \in [k]$, it follows that

$$\begin{aligned}
\mathbb{P}\Big(\mathbb{1}_{E(T^*)}^\mathsf{T} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}}\Big) &= \mathbb{P}\Big(\mathbb{1}_{E(T_i)}^\mathsf{T} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}} \text{ for each } i \in [k]\Big) \\
&= \prod_{i=1}^{k} \mathbb{P}\Big(\mathbb{1}_{E(T_i)}^\mathsf{T} c^* > 2\,\mathrm{OPT}_{\mathrm{HK}}\Big) && \text{by independence of the } k \text{ samplings} \\
&< \left(\frac{1}{2}\right)^k && \text{by Theorem 5.27} \\
&= \frac{1}{2^{\lceil 2\ln n\rceil}} < \frac{1}{2^{2\ln n}} < \frac{1}{2^{\log_2 n}} = \frac{1}{n}\,,
\end{aligned}$$

where the last inequality holds since $\log_2 n = \frac{\ln n}{\ln 2} < 2\ln n$. □

Finally, we join the last results to show we can sample a tree that satisfies the two requirements we established at the beginning of the section.

Theorem 5.29. Consider the context of the statement of Corollary 5.28. Then

$$\mathbb{P}(T^* \text{ is } (\alpha, 2)\text{-thin}) > 1 - \frac{1}{n-1} - \frac{1}{n} > 1 - \frac{2}{n-1}\,,$$

for $\alpha = 4\ln n/\ln\ln n$.

Proof. Denote by A the event that T^* is α-thin, and denote by B the event that $\mathbb{1}_{E(T^*)}^\mathsf{T} c^* \leq 2\,\mathrm{OPT}_{\mathrm{HK}}$. Thus,

$$\mathbb{P}(T^* \text{ is } (\alpha, 2)\text{-thin}) = \mathbb{P}(A \cap B) = 1 - \mathbb{P}\big(\overline{A} \cup \overline{B}\big) \geq 1 - \mathbb{P}\big(\overline{A}\big) - \mathbb{P}\big(\overline{B}\big) > 1 - \frac{1}{n-1} - \frac{1}{n} > 1 - \frac{2}{n-1}\,,$$

where the second equality holds by the probability of the complement (see item (iv) of Proposition 5.1), the first inequality holds by the union bound (see item (ii) from Proposition 5.1), and the second inequality holds by Theorem 5.26 and Corollary 5.28. □

The next result encapsulates Theorem 5.29 so that it can be easily used by the main result of the monograph, Theorem 3.8.

Theorem 5.30. Let D, c, z^*, G_{z^*} be defined as in Definition 3.2. Then there exists a polynomial-time algorithm that finds a $(4\ln n/\ln\ln n, 2)$-thin tree of G_{z^*} with high probability, namely greater than $1-2/(n-1)$.

Proof. By Theorem 5.22, there is a randomized polynomial-time algorithm that samples spanning trees of G_{z^*} using a finite probability space $(\mathcal{T}(G_{z^*}), \mathbb{P})$. Thus, we use this algorithm to sample $\lceil 2\ln n\rceil$ spanning trees of G_{z^*}, and we choose a tree T^* among them that has minimum cost. Both the running time of the algorithm used to sample the spanning trees and the number of times we called this algorithm are polynomial in $|V|$ and $|c|$, the input size of the algorithm ApproxATSP (see Algorithm 3.1); so this whole process takes polynomial time. Moreover, by Theorem 5.29, T^* is a $(4\ln n/\ln\ln n, 2)$-thin tree of G_{z^*} with probability greater than $1 - 2/(n-1)$. □

Chapter 6

The Ellipsoid Method

6.1 The Geometry of Ellipsoids

Throughout this section, we will use V to denote a finite nonempty set.

An **ellipsoid** in $\mathbb{R}^V$ is a set of the form $A\mathbb{B} + b = \{\, Ax + b \in \mathbb{R}^V : x \in \mathbb{B}\}$, where $\mathbb{B}$ is the unit ball in $\mathbb{R}^V$, for an invertible matrix $A \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$.

The next result gives another representation of ellipsoids that uses quadratic forms generated by positive definite matrices. Also, for any ellipsoid, it is shown how to transform its representation with the definition into its representation with quadratic forms and vice-versa.

Proposition 6.1 (Characterization of Ellipsoids)**.** Let E be a subset of $\mathbb{R}^V$. Then E is an ellipsoid if and only if there exists a positive definite matrix $M \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$ such that $E = E(M, b) := \{\, x \in \mathbb{R}^V : (x - b)^\mathsf{T} M^{-1}(x - b) \le 1\}$. Moreover, if $E = A\mathbb{B} + b$ for an invertible matrix $A \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$, then $E = E(AA^\mathsf{T}, b)$; and conversely, if $E = E(M, b)$ for a positive definite matrix $M \in \mathbb{R}^{V\times V}$ and vector $b \in \mathbb{R}^V$, then $E = M^{1/2}\mathbb{B} + b = \{\, M^{1/2}x + b \in \mathbb{R}^V : x \in \mathbb{B}\}$.

Proof. Suppose E is an ellipsoid. Then, by definition, there exists an invertible matrix $A \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$ such that

$$E = \{\, Ay + b \in \mathbb{R}^V : \|y\| \le 1\} = \{\, x \in \mathbb{R}^V : \|A^{-1}(x - b)\| \le 1\}\,.$$

By nonnegativity of norm, for each $x \in E$ we have that

$$\|A^{-1}(x - b)\| \le 1 \Leftrightarrow \|A^{-1}(x - b)\|^2 \le 1^2 \Leftrightarrow (A^{-1}(x - b))^\mathsf{T}(A^{-1}(x - b)) \le 1 \Leftrightarrow (x - b)^\mathsf{T}(AA^\mathsf{T})^{-1}(x - b)$$

Moreover, by Proposition 2.7, the matrix AA^T is positive definite. Thus, we have $E = E(M, b)$ for $M := AA^\mathsf{T}$.

Conversely, suppose $E = E(M, b)$ for a positive definite matrix $M \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$. Recall from item (i) of Theorem 2.6 that $M = M^{1/2}M^{1/2}$ where $M^{1/2}$ is the unique positive definite square root of M. Moreover, by item (ii) of Theorem 2.6, $M^{1/2}$ is invertible, and its inverse is also positive definite. Thus,

$$\begin{aligned} E &= \{\, x \in \mathbb{R}^V : (x - b)^\mathsf{T} M^{-1}(x - b) \le 1\} \\ &= \{\, x \in \mathbb{R}^V : (x - b)^\mathsf{T} M^{-1/2}M^{-1/2}(x - b) \le 1\} \\ &= \{\, x \in \mathbb{R}^V : (M^{-1/2}(x - b))^\mathsf{T} M^{-1/2}(x - b) \le 1\} \\ &= \{\, x \in \mathbb{R}^V : \|M^{-1/2}(x - b)\| \le 1\} \\ &= \{\, M^{1/2}y + b\mathbb{R}^V : y \in \mathbb{B}\}\,. \end{aligned}$$

□

So by Proposition 6.1, a subset E of $\mathbb{R}^V$ is an ellipsoid if there exists a positive definite matrix $M \in \mathbb{R}^{V\times V}$ and a vector $b \in \mathbb{R}^V$ such that $E = E(M, b)$; such vector b is called the **center** of E.

In the Ellipsoid Method, we will produce a sequence of ellipsoids of decreasing volume. In Theorem 6.4, we will show how to produce this sequence, and we will show an upper bound for the ratio of the volumes of consecutive ellipsoids. To do that, we define the volume of a measurable subset of $\mathbb{R}^V$, and we give an equivalent way of writing the definition of an ellipsoid that will help us to calculate its volume.

The **volume** of a measurable subset X of $\mathbb{R}^V$, denoted by $\operatorname{vol}(X)$, is defined by

$$\operatorname{vol}(X) = \int_{x \in X} dx \,. \tag{6.1}$$

An ellipsoid in $\mathbb{R}^V$ is a set of the form $S[\mathbb{B}]$ (see definition of image in Section 2.2) for an invertible affine transformation $S\colon \mathbb{R}^V \to \mathbb{R}^V$, i.e., an ellipsoid is the image of the unit ball by an invertible affine transformation. So with the next lemma, we will be able to compute the volume of an ellipsoid as a scalar of the volume of the unit ball.

Lemma 6.2. Let X be a measurable subset of $\mathbb{R}^V$, and let $S\colon x \in \mathbb{R}^V \mapsto Ax + b$ be an affine transformation for a matrix $A \in \mathbb{R}^{V \times V}$ and a vector $b \in \mathbb{R}$. Then $\operatorname{vol}(S[X]) = |\det(A)| \operatorname{vol}(X)$.

Finally, we introduce another important matrix for our proof of Theorem 6.4, a rotation matrix.

Lemma 6.3. Let x be a vector in $\mathbb{R}^V$, and let $i \in V$. Then there exists a matrix $R \in \mathbb{R}^{V \times V}$, called rotation matrix, such that

$$RR^{\mathsf{T}} = R^{\mathsf{T}}R = I \quad \text{, and} \tag{6.2}$$

$$Rx = \|x\| e_i \,, \tag{6.3}$$

i.e., R is an orthogonal matrix, and it "rotates" x to the multiple vector of e_i of equal norm.

Proof. First, we consider the case where $x = ke_i$ for a real $k \le 0$, i.e., when x is a nonpositive multiple of e_i. Note that

$$\|x\| = \sqrt{k^2} = |k| = -k \,. \tag{6.4}$$

In this case, we have (6.2) and (6.3) for $R \coloneqq -I$. Indeed,

$$(-I)(-I)^{\mathsf{T}} = I = (-I)^{\mathsf{T}}(-I) \quad \text{and} \quad (-I)(ke_i) = -kIe_i = -ke_i \overset{(6.4)}{=} \|x\| e_i \,.$$

Now suppose $x \neq ke_i$ for each real $k \le 0$. Set $y \coloneqq x + \|x\| e_i$, and set $R \coloneqq \frac{2\, yy^{\mathsf{T}}}{\|y\|^2} - I$. Note that $y \neq 0$ since

$$\begin{aligned} x + \|x\| e_i = 0 &\Leftrightarrow x_i = -\|x\| \text{ and } x_j = 0 \text{ for each } j \neq i \Leftrightarrow x_i = -\sqrt{x_i^2} \text{ and } x_j = 0 \text{ for each } j \neq i \\ &\Leftrightarrow -x_i = |x_i| \text{ and } x_j = 0 \text{ for each } j \neq i \Leftrightarrow x_i \le 0 \text{ and } x_j = 0 \text{ for each } j \neq i \\ &\Leftrightarrow x = ke_i \text{ for a real } k \le 0 \,. \end{aligned}$$

So the matrix R can indeed be defined for such x, and then

$$RR^{\mathsf{T}} = R^{\mathsf{T}}R = \left(\frac{2\, yy^{\mathsf{T}}}{\|y\|^2} - I \right)^2 = \frac{4\|y\|^2 yy^{\mathsf{T}}}{\|y\|^4} - \frac{4\, yy^{\mathsf{T}}}{\|y\|^2} + I = I \,,$$

which shows (6.2). Now consider

$$Rx = \frac{2(x + \|x\| e_i)(x^{\mathsf{T}}x + \|x\| e_i^{\mathsf{T}} x)}{\|x + \|x\| e_i\|^2} - x = \frac{2(x + \|x\| e_i)(\|x\|^2 + \|x\| x_i)}{\|x + \|x\| e_i\|^2} - x \,. \tag{6.5}$$

Set

$$\alpha \coloneqq \|x + \|x\| e_i\|^2 = (x + \|x\| e_i)^{\mathsf{T}}(x + \|x\| e_i) = 2(\|x\|^2 + \|x\| x_i) \,, \tag{6.6}$$

and set

$$z \coloneqq 2(x + \|x\| e_i)(\|x\|^2 + \|x\| x_i) \overset{(6.6)}{=} \alpha(x + \|x\| e_i) \,. \tag{6.7}$$

Thus, by rewriting (6.5) with (6.6) and (6.7), we have

$$Rx = \frac{z}{\alpha} - x = \frac{\alpha(x + \|x\| e_i)}{\alpha} - x = \|x\| e_i \,,$$

which shows (6.3). $\square$

The sequence of ellipsoids produced during the Ellipsoid Method has a property: each ellipsoid of the sequence, except the first one, is the ellipsoid of minimum volume that contains a certain "ellipsoidal section" of the previous one in the sequence. We will consider a particular ellipsoidal section that gives rise to one version of the Ellipsoid Method, the **Central-Cut** Ellipsoid Method.

More precisely, for an ellipsoid $E := E(M, b)$ in $\mathbb{R}^V$, consider a halfspace $H := \{\, x \in \mathbb{R}^V : a^\mathsf{T} x \geq a^\mathsf{T} b\}$ in $\mathbb{R}^V$. Set $n := |V|$. Note that the center of E is in the hyperplane determined by $a^\mathsf{T} x = a^\mathsf{T} b$. So E is being "cut" through its center b by the hyperplane determined by $a^\mathsf{T} x = a^\mathsf{T} b$; we call $E \cap H$ a *half-ellipsoid*. The ellipsoid E' of minimum volume that contains the half-ellipsoid $E \cap H$ is the ellipsoid $E(\overline{M}, \overline{b})$ defined by

$$\overline{b} := b + \frac{1}{n+1} \frac{Ma}{\sqrt{a^\mathsf{T} M a}}\,, \tag{6.8a}$$

$$\overline{M} := \frac{n^2}{n^2-1}\left(M - \frac{2}{n+1}\frac{Maa^\mathsf{T} M}{a^\mathsf{T} M a}\right). \tag{6.8b}$$

The ellipsoid E' is called the **Löwner-John ellipsoid** of $E \cap H$. To see the derivation of the above formulas, one can check out Bland, Goldfarb, and Todd [4, Appendix B]. In the next theorem we prove that indeed $E \cap H \subseteq E'$, and we give a bound, smaller than one, for the ratio of the volumes of E' and E. The proof is inspired in Bertsimas and Tsitsiklis [3, Theorem 8.1].

Theorem 6.4. Let $E := E(M, b)$ be an ellipsoid in $\mathbb{R}^V$, and let a be a nonzero vector in $\mathbb{R}^V$. Set $n := |V|$, and set $H := \{\, x \in \mathbb{R}^V : a^\mathsf{T} x \geq a^\mathsf{T} b\}$ to be a halfspace in $\mathbb{R}^V$. Moreover, set $E' := E(\overline{M}, \overline{b})$ to be the Löwner-John ellipsoid of $E \cap H$ as in (6.8). Then

a) $E \cap H \subseteq E'$,

b) $\frac{\operatorname{vol}(E')}{\operatorname{vol}(E)} \leq e^{-1/(2(n+1))} < 1$.

Proof. **a)** First, we consider the particular case where $M = I$, $b = 0$, and $a = e_i$ for an $i \in V$. So $E_0 := E = E(I, 0)$ which is $I^{1/2}\mathbb{B} + 0 = \mathbb{B}$ by Proposition 6.1, i.e., E_0 is the unit ball in $\mathbb{R}^V$. Also, $H_0 := H = \{\, x \in \mathbb{R}^V : x_i \geq 0\}$, and $E_0' := E' = E(\overline{I}, \overline{0})$. In other words, we will show that the ellipsoid E_0' contains the half of the unit ball in $\mathbb{R}^V$ whose points have the i coordinate nonnegative. We start by determining E_0'. By (6.8), we have

$$\overline{0} = 0 + \frac{1}{n+1}\frac{Ie_i}{\sqrt{e_i^\mathsf{T} I e_i}} = \frac{e_i}{n+1}\,, \tag{6.9}$$

$$\overline{I} = \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}\frac{Ie_i e_i^\mathsf{T} I}{e_i^\mathsf{T} I e_i}\right) = \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1} e_i e_i^\mathsf{T}\right). \tag{6.10}$$

Set $u := -\frac{2}{n+1} e_i$, and set $v := e_i$. Then, by the Sherman-Morrison formula (see Proposition 2.5), for the inverse of the second factor in the RHS of (6.10) we have

$$(I + uv^\mathsf{T})^{-1} = I - \frac{uv^\mathsf{T}}{1 + v^\mathsf{T} u} = I - \frac{1}{1 - \frac{2}{n+1} e_i^\mathsf{T} e_i}\left(-\frac{2}{n+1} e_i\right) e_i^\mathsf{T} = I + \frac{n+1}{n-1}\frac{2}{n+1} e_i e_i^\mathsf{T} = I + \frac{2}{n-1} e_i e_i^\mathsf{T}\,,$$

and consequently,

$$\overline{I}^{-1} = \frac{n^2-1}{n^2}\left(I + \frac{2}{n-1} e_i e_i^\mathsf{T}\right). \tag{6.11}$$

Now we find a convenient expression for E_0' to derive the desired property. Since $E_0' = E(\overline{I}, \overline{0})$ and by (6.9) and (6.11), we have

$$
\begin{aligned}
E_0' &= \left\{ x \in \mathbb{R}^V : \left(x - \frac{e_i}{n+1}\right)^\mathsf{T} \left(\frac{n^2-1}{n^2}\right)\left(I + \frac{2}{n-1} e_i e_i^\mathsf{T}\right)\left(x - \frac{e_i}{n+1}\right) \le 1 \right\} \\
&= \left\{ x \in \mathbb{R}^V : \left(\frac{n^2-1}{n^2}\right)\left[\left(x - \frac{e_i}{n+1}\right)^\mathsf{T}\left(x - \frac{e_i}{n+1}\right) + \frac{2}{n-1}\left(x - \frac{e_i}{n+1}\right)^\mathsf{T}(e_i e_i^\mathsf{T})\left(x - \frac{e_i}{n+1}\right)\right] \le 1 \right\} \\
&= \left\{ x \in \mathbb{R}^V : \left(\frac{n^2-1}{n^2}\right)\left[\left(x^\mathsf{T}x - \frac{2x_i}{n+1} + \frac{1}{(n+1)^2}\right) + \frac{2}{n-1}\left(x_i^2 - \frac{2x_i}{n+1} + \frac{1}{(n+1)^2}\right)\right] \le 1 \right\} \\
&= \left\{ x \in \mathbb{R}^V : \left(\frac{n^2-1}{n^2}\right)\left[x^\mathsf{T}x + \frac{2x_i^2}{n-1} + \left(-\frac{2x_i}{n+1} + \frac{1}{(n+1)^2}\right)\left(\frac{n+1}{n-1}\right)\right] \le 1 \right\} \\
&= \left\{ x \in \mathbb{R}^V : \left(\frac{n^2-1}{n^2}\right)\left(x^\mathsf{T}x + \frac{2x_i}{n-1}(x_i - 1) + \frac{1}{n^2-1}\right) \le 1 \right\} \\
&= \left\{ x \in \mathbb{R}^V : \left(\frac{n^2-1}{n^2}\right)x^\mathsf{T}x + \frac{2(n+1)}{n^2}x_i(x_i-1) + \frac{1}{n^2} \le 1 \right\}.
\end{aligned}
\tag{6.12}
$$

Let $x \in E_0 \cap H_0$. So $x \in H_0$ and $x_i \ge 0$. Moreover, by $x_i \ge 0$, the Cauchy-Schwarz inequality, and $x \in E_0 = \mathbb{B}$, we have that $x_i = |e_i^\mathsf{T} x| \le \|e_i\|\|x\| = 1$. Thus, $x_i(x_i - 1) \le 0$. Still by $x \in E_0$, we have that $x^\mathsf{T}x \le 1$. Hence,

$$
\frac{n^2-1}{n^2}x^\mathsf{T}x + \frac{2(n+1)}{n^2}x_i(x_i-1) + \frac{1}{n^2} \le \frac{n^2-1}{n^2} + \frac{1}{n^2} = 1,
$$

i.e., $x \in E_0'$ by the description of E_0' in (6.12). Therefore, $E_0 \cap H_0 \subseteq E_0'$.

Now we consider the general case, as in the statement, where M, b, and a are chosen arbitrarily. We will show an invertible affine transformation $S\colon \mathbb{R}^V \to \mathbb{R}^V$ where (see definitions of image and preimage in Section 2.2)

$$
S[E] = E_0,\ S[H] = H_0 \text{ and } S[E'] = E_0'. \tag{6.13}
$$

Since S is an injective function, by item (i) of Proposition 2.1, we have

$$
S[E] \cap S[H] \subseteq S[E'] \Rightarrow S[E \cap H] \subseteq S[E'] \Rightarrow S^{-1}[S[E \cap H]] \subseteq S^{-1}[S[E']] \Rightarrow E \cap H \subseteq E', \tag{6.14}
$$

where the last implication holds by item (ii) of Proposition 2.1 since S is injective. Finally, since $E_0 \cap H_0 \subseteq E_0'$ and by (6.13) and (6.14), we will have $E \cap H \subseteq E'$. So we build the function S as follows.

Let $T\colon x \in \mathbb{R}^V \mapsto M^{-1/2}(x - b)$. Then $T[E] = M^{-1/2}(M^{1/2}\mathbb{B} + b - b) = \mathbb{B} = E_0$ which satisfies the first condition in (6.13). However, we cannot assert that $T[H] = H_0$ and $T[E'] = E_0'$. So we consider another invertible affine transformation that slightly modifies T. By Lemma 6.3, for any vector $x \in \mathbb{R}^V$ and $j \in V$, there exists a matrix $R \in \mathbb{R}^{V \times V}$, called a rotation matrix, such that $RR^\mathsf{T} = R^\mathsf{T}R = I$ and $Rx = \|x\|e_j$. So define a rotation matrix $R \in \mathbb{R}^{V \times V}$ for $M^{1/2}a$ and i so that

$$
RM^{1/2}a = \|M^{1/2}a\|e_i; \tag{6.15}
$$

also, define the invertible affine transformation $S\colon x \in \mathbb{R}^V \mapsto RM^{-1/2}(x - b)$. We prove S satisfies the conditions in (6.13). First, we show $S[E] = E_0$. By Proposition 6.1, since $E = E(M, b)$, it follows that $E = M^{1/2}\mathbb{B} + b = \{\, M^{1/2}x + b \in \mathbb{R}^V : x \in \mathbb{B}\}$, and so

$$
\begin{aligned}
S[E] &= S[M^{1/2}\mathbb{B} + b] \\
&= \{\, y \in \mathbb{R}^V : \exists x \in \mathbb{B} \text{ s.t. } y = RM^{-1/2}(M^{1/2}x + b - b)\} \\
&= \{\, y \in \mathbb{R}^V : \exists x \in \mathbb{B} \text{ s.t. } y = Rx\} \\
&= \{\, y \in \mathbb{R}^V : y \in \mathbb{B}\} = \mathbb{B},
\end{aligned}
\tag{6.16}
$$

where the second last equality holds since orthogonal matrices preserve norms.

Now we show $S[H] = H_0$. Since $M^{1/2}$ is invertible and $a \neq 0$ by hypothesis, we have $M^{1/2}a \neq 0$. Then $\|M^{1/2}a\| > 0$, and so from (6.15) it follows that

$$
e_i = \tfrac{1}{\|M^{1/2}a\|}RM^{1/2}a. \tag{6.17}
$$

Moreover, by (6.2), note that

$$M^{1/2}R^{\mathsf{T}}RM^{-1/2} = M^{1/2}M^{-1/2} = I\,. \tag{6.18}$$

Thus,

$$\begin{aligned}
y \in H_0 &\Leftrightarrow y \in \mathbb{R}^V \text{ and } e_i^{\mathsf{T}} y \geq 0 \\
&\Leftrightarrow \exists x \in \mathbb{R}^V \text{ s.t. } y = S(x) \text{ and } e_i^{\mathsf{T}} S(x) \geq 0 \\
&\Leftrightarrow \exists x \in \mathbb{R}^V \text{ s.t. } y = S(x) \text{ and } \frac{1}{\|M^{1/2}a\|} a^{\mathsf{T}} M^{1/2} R^{\mathsf{T}} S(x) \geq 0 \qquad \text{by (6.17)} \\
&\Leftrightarrow \exists x \in \mathbb{R}^V \text{ s.t. } y = S(x) \text{ and } a^{\mathsf{T}} M^{1/2} R^{\mathsf{T}} R M^{-1/2}(x-b) \geq 0 \qquad (6.19) \\
&\Leftrightarrow \exists x \in \mathbb{R}^V \text{ s.t. } y = S(x) \text{ and } a^{\mathsf{T}}(x-b) \geq 0 \qquad \text{by (6.18)} \\
&\Leftrightarrow \exists x \in \mathbb{R}^V \text{ s.t. } y = S(x) \text{ and } x \in H \\
&\Leftrightarrow y \in S[H]\,,
\end{aligned}$$

where the "if" part of the second "if and only if" of (6.19) holds since S is surjective.

Finally, we show $S[E'] = E_0'$. Again by Proposition 6.1, since $E' = E(\overline{M}, \bar{b})$, it follows that $E' = \overline{M}^{1/2}\mathbb{B} + \bar{b}$, and so

$$\begin{aligned}
S[E'] &= RM^{-1/2}\Big(\overline{M}^{1/2}\mathbb{B} + \bar{b} - b\Big) \overset{(6.8a)}{=} RM^{-1/2}\left(\overline{M}^{1/2}\mathbb{B} + \frac{1}{n+1}\frac{Ma}{\sqrt{a^{\mathsf{T}}Ma}}\right) \\
&= RM^{-1/2}\overline{M}^{1/2}\mathbb{B} + \frac{1}{n+1}\frac{RM^{1/2}a}{\|M^{1/2}a\|} \overset{(6.15)}{=} RM^{-1/2}\overline{M}^{1/2}\mathbb{B} + \frac{1}{n+1}\frac{\|M^{1/2}a\|e_i}{\|M^{1/2}a\|} \qquad (6.20) \\
&\overset{(6.9)}{=} RM^{-1/2}\overline{M}^{1/2}\mathbb{B} + \bar{0}\,.
\end{aligned}$$

In the RHS of the last equation, each matrix in the product $RM^{-1/2}\overline{M}^{1/2}$ is invertible, so the product also is. Thus, $S[E']$ is an ellipsoid whence, by Proposition 6.1, $S[E'] = E(D, \bar{0})$ with $D := (RM^{-1/2}\overline{M}^{1/2})(RM^{-1/2}\overline{M}^{1/2})^{\mathsf{T}}$. Since $E_0' = E(\bar{I}, \bar{0})$, it remains to show that $D = \bar{I}$. Indeed,

$$\begin{aligned}
D &= RM^{-1/2}\overline{M}M^{-1/2}R^{\mathsf{T}} \overset{(6.8b)}{=} \frac{n^2}{n^2-1}\left(RM^{-1/2}MM^{-1/2}R^{\mathsf{T}} - \frac{2}{n+1}\frac{RM^{1/2}aa^{\mathsf{T}}M^{1/2}R^{\mathsf{T}}}{a^{\mathsf{T}}Ma}\right) \\
&\overset{(6.2)}{\underset{(6.15)}{=}} \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}\frac{\|M^{1/2}a\|^2 e_i e_i^{\mathsf{T}}}{\|M^{1/2}a\|^2}\right) = \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}e_i e_i^{\mathsf{T}}\right) \overset{(6.10)}{=} \bar{I}\,.
\end{aligned}$$

b) Since $E_0' = E(\bar{I}, \bar{0})$, by Proposition 6.1, we have that $E_0' = \bar{I}^{1/2}\mathbb{B} + \bar{0} = \bar{I}^{1/2}E_0 + \bar{0}$. Then, by Lemma 6.2, $\operatorname{vol}(E_0') = |\det(\bar{I}^{1/2})| \operatorname{vol}(E_0)$. Moreover, by Lemma 6.2 again, we have that

$$\begin{aligned}
\operatorname{vol}(E') &= \operatorname{vol}(S^{-1}[S[E']]) = k\operatorname{vol}(S[E']) = k\operatorname{vol}(E_0') \quad \text{, and} \\
\operatorname{vol}(E) &= \operatorname{vol}(S^{-1}[S[E]]) = k\operatorname{vol}(S[E]) = k\operatorname{vol}(E_0)\,,
\end{aligned}$$

for some $k \in \mathbb{R}_+$. Thus,

$$\frac{\operatorname{vol}(E')}{\operatorname{vol}(E)} = \frac{\operatorname{vol}(E_0')}{\operatorname{vol}(E_0)} = |\det(\bar{I}^{1/2})|\,. \tag{6.21}$$

Since $\det(AB) = \det(A)\det(B)$ for any matrices $A, B \in \mathbb{R}^{V\times V}$, we have $\det(\bar{I}) = \det(\bar{I}^{1/2})^2$, whence

$|\det(\overline{I}^{1/2})| = \sqrt{\det(\overline{I})}$ and

$$
\begin{aligned}
\sqrt{\det(\overline{I})} &\overset{(6.10)}{=} \left[\det\left(\frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}e_ie_i^\mathsf{T}\right)\right)\right]^{1/2} \\
&= \left(\frac{n^2}{n^2-1}\right)^{n/2}\left[\det\left(I - \frac{2}{n+1}e_ie_i^\mathsf{T}\right)\right]^{1/2} && c\det(A) = c^n\det(A) \text{ for any } A \in \mathbb{R}^{V\times V}, c \in \mathbb{R} \\
&= \left(\frac{n^2}{n^2-1}\right)^{n/2}\left(1 - \frac{2}{n+1}\right)^{1/2} && \text{determinant of a diagonal matrix} \\
&= \left(\frac{n^2}{n^2-1}\right)^{(n-1)/2}\left(\frac{n^2}{n^2-1}\right)^{1/2}\left(\frac{n-1}{n+1}\right)^{1/2} \\
&= \left(\frac{n^2}{n^2-1}\right)^{(n-1)/2}\left(\frac{n^2}{(n+1)^2}\right)^{1/2} \\
&= \frac{n}{n+1}\left(\frac{n^2}{n^2-1}\right)^{(n-1)/2} \\
&= \left(1 - \frac{1}{n+1}\right)\left(1 + \frac{1}{n^2-1}\right)^{(n-1)/2} \\
&\leq e^{-1/(n+1)}e^{(n-1)/(2(n^2-1))} && e^x \geq 1 + x \text{ for each } x \in \mathbb{R} \text{ (see Proposition 2.3)} \\
&= e^{-1/(n+1)}e^{1/(2(n+1))} \\
&= e^{-1/(2(n+1))}.
\end{aligned}
$$

□

6.2 The Central-Cut Ellipsoid Method

Throughtout this section, we use U and V to denote finite noempty sets. Also, set $n := |V|$, and set $m := |U|$.

Our problem is, given a system of linear inequalities $Ax \leq b$, for a matrix $A \in \mathbb{R}^{U\times V}$ and a vector $b \in \mathbb{R}^U$, that determines a polyhedron P, decide whether P is empty or not, and if P is nonempty, we want to determine a point in it. The ellipsoid method is an algorithm that solves this problem efficiently. We will present this algorithm with three assumptions: two about the polyhedron P and one about the precision of the arithmetic operations that the algorithm performs. One can show how to reduce any instance of the problem to one with these two assumptions about the polyhedron, and one can show how to deal with the precision issues that arise from the arithmetic operations used; however, we will not cover this here. For a complete treatment, one can look at Grötschel, Lovász, and Schrijver [12, Chapter 3].

In essence, the algorithm produces a sequence of ellipsoids of decreasing volume that contain the polyhedron P. This implies our first assumption about P: it must be bounded. Moreover, we require that the first ellipsoid be a ball $B(z,r)$ of positive volume v_0 (It is easy to see that $B(z,r) = rI\mathbb{B} + z = E(r^2I, z)$, i.e., $B(z,r)$ is indeed an ellipsoid).

Now, for each ellipsoid $E = E(M,z)$ produced during its execution, the algorithm tests whether the center z of E lies in P. If $z \in P$, the algorithm returns z and terminates. Otherwise, there is a valid inequality for P that is violated by z, and then with the halfspace H determined by this inequality, the algorithms builds the next ellipsoid E' of the sequence. From formulas in (6.8), the method set E' to be the Löwner-John ellipsoid of the half-ellipsoid $E \cap H$. By Theorem 6.4, the ellipsoid E' will contain $E \cap H$, and it will have smaller volume than E. Also, as both E and H contain P, the ellipsoid E' will also contain P.

Then comes the second assumption concerning how the algorithm decides to stop the search for a point in P and so asserts that P is empty. The algorithm assumes that either P is empty or $\operatorname{vol}(P) > v$ for some given $v \in \mathbb{R}_{++}$. Moreover, as the algorithm advances, the volume of the ellipsoids decreases. Thus, as it will be shown in Theorem 6.5, in at most k iterations (a number in function of n, v_0, and v), either a point of P will be found and returned or an ellipsoid of volume smaller than or equal to v, containing P, will be generated, and the polyhedron will be asserted to be empty.

Finally, to generate the ellipsoids in the algorithm, some computations such as taking square roots are required. So our algorithm assumes that any arithmetic operation can be computed with infinite precision and in unit time. That said, a more precise description of the algorithm is as follows, ignoring some implementation details. We call it the **central-cut** ellipsoid method since, as we discussed before the Theorem 6.4, we build a sequence of ellipsoids with the Löwner-John ellipsoids of half-ellipsoids. The next algorithm description and Theorem 6.5 are inspired in Bertsimas and Tsitsiklis [3, Section 8.3].

Algorithm 6.1: CentralCutEllipsoidMethod(A, b, z, r, v)

Input:

(i) A matrix $A \in \mathbb{R}^{U \times V}$ and a vector $b \in \mathbb{R}^U$, where $n := |V|$ and $m := |U|$, that determines a polyhedron P.

(ii) A vector $z \in \mathbb{R}^V$ and a scalar $r \in \mathbb{R}_{++}$ s.t. the ball $B(z, r)$, of volume v_0, contains P.

(iii) A scalar $v \in \mathbb{R}_{++}$ s.t. $\operatorname{vol}(P) > v$ if P is nonempty, and $v < v_0$.

Output: Either assert P is empty or return a point in P

$k := \lceil 2(n+1)\log(v_0/v) \rceil$
$E_0 := E(M_0, z_0)$, where $M_0 := r^2 I$ and $z_0 := z$, i.e., E_0 is the ball $B(z, r)$
for $i \leftarrow 0$ **to** $k-1$ **do**
    if $z_i \in P$ **then**
        return z_i
    else
        Find a row ℓ_i of A s.t. $a_{\ell_i}^\mathsf{T} z_i > b_{\ell_i}$ for $a_{\ell_i} := (e_{\ell_i}^\mathsf{T} A)^\mathsf{T}$, i.e., a valid inequality for P violated by z_i
        Define M_{i+1}, z_{i+1} and so the ellipsoid $E_{i+1} := E(M_{i+1}, z_{i+1})$ by applying the formulas (6.8a) and (6.8b) to M_i, z_i, and a_{ℓ_i} so that $E_i \cap H_i \subseteq E_{i+1}$, where $H_i := \{\, x \in \mathbb{R}^V : a_{\ell_i}^\mathsf{T} x \le a_{\ell_i}^\mathsf{T} z_i \}$ is a halfspace
return P is empty

Theorem 6.5 (Correctness of Ellipsoid Method)**.** Let $A \in \mathbb{R}^{U \times V}$ be a matrix, and let $b \in \mathbb{R}^U$. Set $P := \{\, x \in \mathbb{R}^V : Ax \le b \}$ to be a polyhedron. Suppose that P is bounded and that P either is empty or has positive volume. Let $z \in \mathbb{R}^V$ and let $r \in \mathbb{R}_{++}$ such that $P \subseteq B(z, r)$, and set $v_0 := \operatorname{vol}(B(z, r))$. Let $v \in \mathbb{R}_{++}$ be such that $\operatorname{vol}(P) > v$ if P is nonempty. Set $k := \lceil 2(n+1)\log(v_0/v) \rceil \ge 1$. Set $M_0 := r^2 I$ and set $z_0 := z$ so that $E_0 := E(M_0, z_0) = B(z, r)$. For each $i \in \{0, \dots, k-1\}$, if z_i is defined and there exists $\ell_i \in [m]$ such that $a_{\ell_i}^\mathsf{T} z_i > b_{\ell_i}$ for $a_{\ell_i} := (e_{\ell_i}^\mathsf{T} A)^\mathsf{T}$, then define M_{i+1} and z_{i+1} by applying the formulas (6.8a) and (6.8b) to M_i, z_i, and a_{ℓ_i}, and set $E_{i+1} := E(M_{i+1}, z_{i+1})$. Moreover, set j to be the largest integer in $\{0, \dots, k\}$ such that E_j is defined. Then either $j < k$ and $z_j \in P$, or $j = k$ and P is empty.

Proof. If $j < k$, then $Az_j \le b$, i.e., $z_j \in P$. Suppose $j = k$. We want to show that P is empty which, by our assumption on P, is equivalent to show that $\operatorname{vol}(P) \le v$. We will show $\operatorname{vol}(P) \le v$ by showing that $P \subseteq E_k$ and that $\operatorname{vol}(E_k) \le v$.

First, we show, using induction on $i \in \{0, 1, \dots, k\}$, something stronger than $P \subseteq E_k$, that is, $P \subseteq E_i$ for each $i \in \{0, 1, \dots, k\}$. By hypothesis, we have $P \subseteq E_0$. Let $0 \le i < k$, and suppose $P \subseteq E_i$. We will show that $P \subseteq E_{i+1}$. Since $i < k = j$, we have $a_{\ell_i}^\mathsf{T} z_i > b_{\ell_i}$. Set $H_i := \{\, x \in \mathbb{R}^V : a_{\ell_i}^\mathsf{T} x \le a_{\ell_i}^\mathsf{T} z_i \}$ to be a halfspace. Then $P \subseteq H_i$. Moreover, by induction hypothesis, $P \subseteq E_i$. Therefore, we have $P \subseteq E_i \cap H_i$, and then, by item a) of Theorem 6.4, it follows that $P \subseteq E_{i+1}$. Thus, $P \subseteq E_k$.

Now, by induction on $i \in \{0, 1, \dots, k\}$ and using item b) of Theorem 6.4, we have that

$$\operatorname{vol}(E_k) < \operatorname{vol}(E_0) e^{-k/(2(n+1))} = v_0 e^{-\lceil 2(n+1)\log(v_0/v) \rceil/(2(n+1))} \le v_0 e^{-\log(v_0/v)} = v\,. \qquad \square$$

A few observations are worth making. First, we remark on the assumption of $v < v_0$ that implies $k \ge 1$. If $v \ge v_0$, then, by the assumptions in the input of the algorithm, we would have that P is empty and the algorithm would not be necessary. Second, we remark on the value of k. We want k to be the smallest positive integer for which $\operatorname{vol}(E_k) \le v$ so that, if the algorithm reaches this point, we can assert that P is empty. Since we have used formulas (6.8a) and (6.8b) to build the sequence of ellipsoids in the algorithm, we used the bound in item b) of Theorem 6.4 to determine such k. Thus, finding such k reduces to solving $v_0 e^{-k/2(n+1)} \le v$.

6.3 Equivalence of Separation and Optimization

In this section, most definitions and the statement of Theorem 6.8 are from Schrijver [15, Section 5.11.]. We present them, with a slight rewriting in some parts, so that we can use Theorem 6.8 in the proof of Theorem 3.5 — the result that shows we can perform Line 1 of ApproxATSP (see Algorithm 3.1) in polynomial-time.

Let Σ be a finite set, called **alphabet**, whose elements are called **symbols**. Then set Σ^* to be the set of all finite sequences of symbols, called **words** or **strings**, of Σ, that is, $\Sigma^* := \{\,(w_1,\dots,w_k) : k \in \mathbb{N}, w \in \Sigma^k\}$. For any $k \in \mathbb{N}$, abbreviate words $w = (w_1,\dots,w_k)$ by $w = w_1 \dots w_k$. The **size** of a word w in Σ^*, denoted by $|w|$, is the number of symbols in the sequence that defines w. Let Π be a subset Σ^*, called a **language** over Σ. Then define the family $(P_\sigma)_{\sigma\in\Pi}$ where P_σ is a rational polyhedron in $\mathbb{Q}^{E_\sigma}$ and E_σ is a finite set. We state two assumptions that a family such as $(P_\sigma)_{\sigma\in\Pi}$ can have:

There is a polynomial-time algorithm that, given $\sigma \in \Sigma^*$, tests if σ belongs to Π and, if so, returns the set E_σ; (6.22a)

There is a polynomial p such that P_σ is determined by linear inequalities each of size at most $p(\text{size}(\sigma))$. (6.22b)

If P is a nonempty polyhedron in $\mathbb{R}^V$, for a finite and nonempty set V, there exists a cone charcone(P) associated to it, called **characteristic cone** of P, defined by:

$$\text{charcone}(P) := \{\, y \in \mathbb{R}^V : x + \lambda y \in P \text{ for each } x \in P,\ \lambda \geq 0\}\,, \tag{6.23}$$

Now we define the separation and optimization problems, and then Theorem 6.8 presents how they are related with respect to being solvable in polynomial-time.

Definition 6.6. The **optimization problem** for $(P_\sigma)_{\sigma\in\Pi}$ is the problem:
Input: $\sigma \in \Pi$ and $c \in \mathbb{Q}^{E_\sigma}$
Task: solve $\max\{\, c^\mathsf{T} x : x \in P_\sigma\}$ which can be formulated as an LP, and so, by Theorem 2.4, have three possible solutions: the LP is unfeasible, i.e., P_σ is empty; the LP is unbounded, i.e., there is $y \in \text{charcone}(P_\sigma)$ with $c^\mathsf{T} y > 0$; the LP has an optimum solution, i.e., there is $x \in P_\sigma$ maximizing $c^\mathsf{T} x$ over P_σ.

Definition 6.7. The **separation problem** for $(P_\sigma)_{\sigma\in\Pi}$ is the problem:
Input: $\sigma \in \Pi$ and $\overline{x} \in \mathbb{Q}^{E_\sigma}$
Task: decide if $\overline{x}$ belongs to P_σ, and if not, find a separating hyperplane between $\overline{x}$ and P_σ, that is, find a vector $c \in \mathbb{Q}^{E_\sigma}$ such that $c^\mathsf{T} x < c^\mathsf{T}\overline{x}$ for each $x \in \mathbb{Q}^{E_\sigma}$.

Theorem 6.8 ([15, Theorem 5.10.])**.** Let $\Pi \subseteq \Sigma^*$, and let $(P_\sigma)_{\sigma\in\Pi}$ satisfy (6.22). Then the optimization problem for $(P_\sigma)_{\sigma\in\Pi}$ is polynomial-time solvable if and only if the separation problem for $(P_\sigma)_{\sigma\in\Pi}$ is polynomial-time solvable.

References

[1] N. Anari and S. O. Gharan. "Effective-resistance-reducing flows, spectrally thin trees, and asymmetric TSP". In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015*. IEEE Computer Soc., Los Alamitos, CA, 2015, pages 20–39. URL: https://doi.org/10.1109/FOCS.2015.11 (cited on page 1).

[2] A. Asadpour, M. X. Goemans, A. Madry, 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 pages i, 1, 2, 12).

[3] D. Bertsimas and J. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific, 1997 (cited on pages 57, 61).

[4] R. G. Bland, D. Goldfarb, and M. J. Todd. "The ellipsoid method: a survey". In: *Oper. Res.* 29.6 (1981), pages 1039–1091. URL: https://doi.org/10.1287/opre.29.6.1039 (cited on page 57).

[5] C. Chekuri, J. Vondrak, and R. Zenklusen. "Dependent Randomized Rounding for Matroid Polytopes and Applications". In: (2009). URL: http://arxiv.org/abs/0909.4348 (cited on page 38).

[6] N. Christofides. *Worst-case analysis of a new heuristic for the travelling salesman problem*. Technical Report 388. Graduate School of Industrial Administration, Carnegie Mellon University, 1976 (cited on page 1).

[7] W. J. Cook. *In pursuit of the traveling salesman*. Mathematics at the limits of computation. Princeton University Press, Princeton, NJ, 2012, pages xvi+228 (cited on page 1).

[8] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver. *Combinatorial optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1998, pages x+355.

[9] A. M. Frieze, G. Galbiati, and F. Maffioli. "On the worst-case performance of some algorithms for the asymmetric traveling salesman problem". In: *Networks* 12.1 (1982), pages 23–39 (cited on page 1).

[10] M. R. Garey and D. S. Johnson. *Computers and intractability*. A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences. W. H. Freeman and Co., San Francisco, Calif., 1979, pages x+338 (cited on page 13).

[11] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Third. Oxford University Press, New York, 2001, pages xii+596 (cited on page 30).

[12] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*. Second. Volume 2. Algorithms and Combinatorics. Springer-Verlag, Berlin, 1993, pages xii+362. URL: https://doi.org/10.1007/978-3-642-78240-4 (cited on page 60).

[13] P. R. Halmos. *Naive set theory*. Reprint of the 1960 edition, Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1974, pages vii+104 (cited on page 50).

[14] D. R. Karger. "Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm". In: *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, TX, 1993)*. ACM, New York, 1993, pages 21–30 (cited on pages 46, 49).

[15] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. A*. Volume 24. Algorithms and Combinatorics. Paths, flows, matchings, Chapters 1–38. Springer-Verlag, Berlin, 2003, pages xxxviii+647 (cited on pages 16, 24, 62).

[16] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. B*. Volume 24. Algorithms and Combinatorics. Matroids, trees, stable sets, Chapters 39–69. Springer-Verlag, Berlin, 2003, i–xxxiv and 649–1217 (cited on pages 38–40, 45).

[17] A. Schrijver. *Theory of linear and integer programming*. Wiley-Interscience Series in Discrete Mathematics. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, 1986, pages xii+471 (cited on page 40).

[18] O. Svensson, J. Tarnawski, and L. A. Végh. "A constant-factor approximation algorithm for the asymmetric traveling salesman problem". In: *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2018, pages 204–213. URL: https://doi.org/10.1145/3188745.3188824 (cited on page 2).

[19] V. Traub and J. Vygen. "An improved approximation algorithm for ATSP". In: *CoRR* abs/1912.00670 (2019). URL: http://arxiv.org/abs/1912.00670 (cited on page 2).

[20] V. V. Vazirani. *Approximation algorithms*. Springer-Verlag, Berlin, 2001, pages xx+378 (cited on page 1).