SOME PROBLEMS ON DISCRETE GEOMETRY AND COMBINATORICS

BY LEI WANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

William Steiger and Mario Szegedy

and approved by

New Brunswick, New Jersey

JANUARY, 2012

ABSTRACT OF THE DISSERTATION

Some Problems On Discrete Geometry and Combinatorics

by Lei Wang

Dissertation Directors: William Steiger and Mario Szegedy

Let K be a convex body in the plane. It is known that K can never be partitioned into seven regions of equal area by three non-concurrent lines. We will be concerned with a partition of K by three non-concurrent lines such that the ratio of the area of smallest region to the area of biggest region is maximum. We call this an optimal balanced partition at K. We show that the best possible ratio is achieved when K is a triangle and we characterize the optimal balanced partition in this case. We conjecture that the condition holds for optimal balanced partitions of all convex bodies but can only prove a weaker result.

In the second part of the thesis, we switch to the zigzag problem. We are given a set of n points in $\mathbb{R}^2$ and seek the minimum number of line segments required for a polygonal chain (or a simple polygonal chain) to traverse all the points. We show an $n/2 + O(n/\log n)$ upper bound if self-intersection is allowed and an $n - \lfloor \frac{n-2}{8} \rfloor$ upper bound if self-intersection is not allowed.

The third part of this thesis is about finding the optimally balanced forward degree sequence of a graph. The final part studies the optimal solutions for some variants of the Towers of Hanoi problem.

Acknowledgements

First and foremost I want to thanks my advisors Dr. Mario Szegedy and Dr. William Steiger. It has been an honor to be their student. I appreciate all their contributions of time, ideas and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm they have for their research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. Without their encouragement and effort, this thesis would not have been completed or written.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. I am especially grateful for Xiaomin Chen and Bin Tian, who have contributed immensely to my personal and professional time at Rutgers. We worked together on the Hanoi Tower problem, and I very much appreciated their enthusiasm, intensity, willingness to solve hard theoretic problems.

For this dissertation I would like to thanks my reading committee members: Dr. Mario Szegedy, Dr. William Steiger, Dr. Bahman Kalantari and Dr. Boris Aronov for their time, interest, and helpful comments. I would also like to thank our graduate secretary Carol DiFrancesco who was so patient and helped me through all those complicated paper works.

My time at Rutgers was made enjoyable in large part due to the many friends that became a part of my life. I am grateful for time spent with roomates and friends, for Xiaomin and Jun, Bin and Liming, Fengming and Ye, Ming and Xiaotao, Ruoming and Zhiyan and their lovely son Amao. The time we have together became my most precious memories that will never fade away.

Lastly, I would like to thank my family for all their love and encouragement: for my sister Ying Wang and brother-in-law Shuai Jing who encouraged me during the toughest times; for my new born nephew, cute and smiley Momo; and most of all for my parents who raised me with a love of science and supported me in all my pursuits.

Thank you.

Dedication

For my family, who offered me unconditional love and support throughout the course of this thesis.

Table of Contents

Chapter 1

Introduction

Discrete and computational geometry is a vigorous subject that has grown out of Euclidean geometry and then was greatly stimulated by questions and methods from computer science and combinatorics. It is also driven by practical, real-world problems of a geometric nature and algorithmic and computational questions that have then originated in a natural way. This thesis touches a variety of different problems, but all relate to ideas from discrete and computational geometry in a natural way.

In 1949 Buck and Buck [2] showed that for any convex body $K \subset \mathbb{R}^2$ whose area is 1, there always exists three concurrent lines l_0, l_1, l_2 that equipartition K, i.e., divides K into six regions of area $1/6$ each. They also showed that no convex body K can be partitioned by three lines lines l_0, l_1, l_2 into seven regions of area $1/7$ each.

Since a convex body K can never be equipartitioned into seven regions, Buck and Buck raised the following question: can K be partitioned into seven regions such that there is one of area $t \in (0, 1/7) \cup (1/7, 1)$, and six of area $(1-t)/6$ each? They showed that if such a partition exists, it must be the central region that has area t. Hence, such a partition received its name as *six-way equipartition*. Finally, Buck and Buck conjectured that if a six-way equipartition exists, then $t \leq 1/49$, which may be achieved when K is a triangle.

This conjecture was proved a year later by Sholander [4]. He showed that the triangle is the extreme convex body which maximize the area of central region in a six-way equipartition, if such a partition exists. Later, there has followed a succession of interesting results [8], [9], [10], [11], [13], [14], [15], [16] about the possibilities/impossibilities of partitioning measures in various ways.

Surprisingly, the main existence question for six-way equipartitions of planar convex bodies remained open for sixty years! In 2010, Steiger, Szegedy and Zhao [6] showed that, given a convex body $K \subset \mathbb{R}^2$ and a unit vector $v \in \mathbb{R}^2$, there exists a unique trio of lines that form a six-way equipartition of K, with one of them having normal vector v.

Recently, Steiger, Szegedy, Wang and Kalantari [7] addressed the following question which had been posed by Kalantari: instead of looking for six equal areas and one different area, what if we seek a partition of the convex body K by three lines such that the ratio of the smallest area to the biggest area is maximized. We name such a partition as *the most balanced partition,* using the ratio as the measure of balance. If three lines are concurrent, K is partitioned into six regions and the balance ratio is 0 (the point of concurrency is a triangle of area 0). If three lines are not concurrent, K is partitioned into seven regions and we know from Buck and Buck [2] that optimal ratio 1 can never be achieved. In the latter case, seeking the most balanced partition for a convex body K and seeking the extreme convex body that achieves the best ratio are interesting problems and we study them in Chapter 2.

Another interesting, though hard, partitioning problem is to simultaneously equipartition multiple point sets in the plane. The Ham-Sandwich Theorem tells us that two point sets in the plane can always be simultaneously equipartitioned by a straight line. Futhermore, Bárány and Matousek [10] showed that, there always exists a *wedge* that simultaneously equipartitions three point sets in the plane. Their proof is purely topological. Later, Bereg [11] gave a combinatoric proof which led to an efficient algorithm that constructs the desired wedge.

A natural extension of a line or a wedge with number of turns 1, is a polygonal path with number of turns k. Clearly, polygonal paths with more turns may simultaneously equipartition point sets with more different types of points. It turns out to be very hard to determine the minimun number of turns k required to simultaneously equipartition any n sets of points in the plane.

A huge simplification of the above problem is to make each point set a simple point. Thus, we are looking for the minimum number of segments that traverses n points in the plane, which we call the *zigzag* problem. Though a simplified version of the zigzag problem has been studied in [17], [18], where only axis-aligned segments are allowed, nothing better than trivial bounds is known for the general case. In chapter 3, we discuss zigzag problem and improve its upper bound.

Here is the classical secretary problem, which is well studied: There are n applicants arriving (online) one by one. Each has value v_i and the goal is to hire the best. After interviewing applicant i, we learned his value v_i. At this point we either reject him and continue interviewing more candidates, or hire him, and interviewing stops.

Here, we study a problem with the same flavor, which is called the *offer rejection problem*:

- There are n almost equally good candidates on the waiting list.
- Some pairs of the candidates can work together, some pairs cannot. These pairs are known in advance.
- After giving an offer to candidtate i, there is a probability p_i candidate i will take the offer. Again, all p_i's are known in advance.

We will hire exactly two candidates. The goal is to select the best order to give the offers to maximize the likelihood of hiring two candidates that can work together.

The offer rejection problem leads naturally to a subject that is interesting in its own right, namely, the *forward degree sequences of graphs*. A forward degree sequence arises from an ordering σ of the vertices of a graph. We eliminate the vertices according to this ordering, and the forward degree d_v^σ of a vertex v is its degree in the remaining graph when we eliminate it.

The idea of forward degree sequences is related to two classical topics in graph theory, namely, the *degree sequence* and *vertex elimination order*. The degree sequences of graphs are well characterized in [20], [21], and [22]. The vertex elimination order gives a nice characterization of chordal graphs (see [25]). The forward degree sequences we define and study here arise from very different questions and are of different nature.

Several nice connections between the offer rejection problem and the forward degree sequences are known. Here we focus on the pure graph-theoretical aspects. One connection arises when we associate to each forward degree sequence a polynomial $P_\sigma(z) = \sum_{v\in V} z^{d_v^\sigma}$. We find that the offer rejection problem with rejection probability q is equivalent to the problem of finding the σ which minimizes $P_\sigma(1/q)$ in the graph where edges represent pairs that can not work together. We will define an ordering σ to be more *balanced* than τ if $P_\sigma(1/q) \le P_\tau(1/q)$ for every probability q. A related notion is *strongly balanced*. It is an interesting combinatorial question whether a graph has a most balanced (strongly balanced) forward degree sequence. We prove that this is true in some nice classes of graphs, for example, chordal graphs and 3-regular graphs, where we also give a polynomial time algorithm to find the most balanced sequence.

The forward degree sequences carry a lot of information about their graphs. One may easily express some usual graph parameters in terms of properties of the forward degree sequences (see Section 4.2). Here we define some new graph parameters (Section 4.3.3) based on the forward degree sequences. These parameters, besides their close relation to the offer rejection problem, are of purely graph-theoretic interest as well. One of the interesting problems that remains open is how to compute some of the parameters in polynomial time.

We study forward degree sequence in chapter 4 and we show that every 3-regular graph has a most balanced forward degree sequence. This gives some new insight into the graph isomorphism problem discussed in [23] and [24].

In chapter 5, we discuss two new variations of the famous Towers of Hanoi problem. The Towers of Hanoi was invented by the French mathematician Edouard Lucas in 1883. In the legend, 64 sacred disks were initially stacked in increasing size on one of three pegs, with the largest at the bottom. A monk has to move the entire tower to another peg. The disks are fragile; only one can be carried at a time. And most importantly, the monk must obey the following rule:

The Divine Rule: A disk may not be placed on top of a smaller disk.

The disks are of size $1, 2, \ldots, n$, usually we denote the disk of size i by $\langle i\rangle$; and we denote the set of disks by $[n] = \{\langle i\rangle : 1 \le i \le n\}$. The three pegs are P_1, P_2, and P_3.

The Towers of Hanoi and its many variations are well studied. Here we just mention a few of them: The cyclic towers of Hanoi was first studied by Atkinson [26]. In [32], Klein and Minsker solved the variation where there can be bigger disks above smaller disks in the initial configuration, but the moves still obey the divine rule. The multi-peg Towers of Hanoi problem was proposed by Stewart [34], and remains a big open problem (see related works by Stewart [35], Frame [30], Szegedy [37], Klavžar, Milutinović, and Petr [31], and Chen and Shen [27]). For a good bibliography with more than 200 entries on this subject, see Paul Stockmeyer's manuscript ([36]).

In chapter 5 we study two new variations of the game. Both are on three pegs, and the monk may violate the divine rule slightly in some way. We give procedures to solve these two versions, and prove the optimality of our procedures.

Chapter 2

Optimal Partitioning by Three Lines in the Plane

2.1 Introduction

It is well known that a line can be drawn through any interior point of a convex body K in the plane, partitioning it into two pieces of equal area. If we consider two lines which intersect in K, they partition it into four regions. It is an easy consequence of the Ham-Sandwich Theorem that there always exist two lines partitioning a convex body into four regions of equal area. Furthermore, Courant and Robbins [3] showed that those two lines can be chosen to be perpendicular.

When it comes to three lines that have all pairwise intersections in K, there are two cases: if the three lines are concurrent, they partition K into six regions; otherwise, they partition K into seven regions. Let us assume $|K| = 1$, where we write $|S|$ for the area of set $S \subset \mathbb{R}^2$. Buck and Buck [2] showed that for any convex body K in the plane, there always exists three concurrent lines that equipartition K into six regions, i.e., each region has area $1/6$. They also showed that no convex body K can be partitioned by three non-concurrent lines into seven regions, each of area $1/7$.

Buck and Buck further proved that if there is a partition where six of the seven regions have equal area (this is called a six-way equiparition), these six regions are necessarily the six outer regions. They conjectured that, in this case, the area of the inner region is at most one eighth the area of the outer regions, i.e., at most $1/49$. This value is achieved when K is a triangle and K is six-way equipartitioned by three lines parallel to its own sides, a partition easily seen to exist. Buck and Buck's conjecture was later proved by Sholander [4], who showed that the triangle is the extreme convex body admitting a six-way equipartitioning: for any convex body K that has a six-way equipartition, if K is not a triangle, then the area of the central triangle is less than

1/49.

The general existence question for six-way equipartitions of convex bodies in the plane has remained open for a long time, until Steiger, Szegedy and Zhao [6] settled it by showing for any convex body, six-way equipartitions always exist and, if we fix the direction of one line, are unique.

2.2 Most balanced partition

Consider a convex body K with $|K| = 1$ and lines ℓ_0, ℓ_1, ℓ_2, each pair of which intersects in K. They form seven regions $A_1, A_2, \cdots, A_7$, as in Figure 2.1. Let $m = min_{1 \leq i \leq 7}\{|A_i|\}$ and $M = max_{1 \leq i \leq 7}\{|A_i|\}$. By Buck and Buck's result [2], we know that $m < M$. Our goal is to maximize m/M, i.e., to find the most balanced partition.

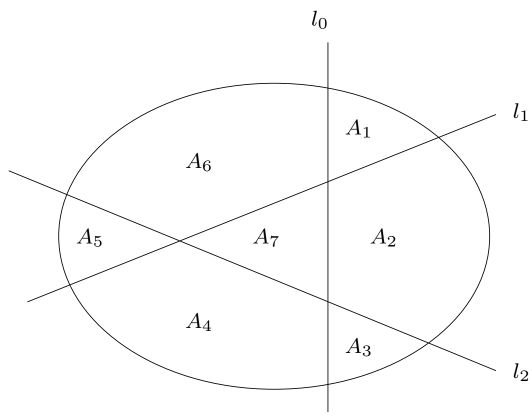

Figure 2.1: lines ℓ_0, ℓ_1, ℓ_2 partition K into seven regions $A_1, A_2, \cdots, A_7$

There is a nice class of partitions of our special interest, which we call balanced partitions.

Definition 1: *ℓ_0, ℓ_1, ℓ_2 form a balanced partition of K if*

1. $|A_1| = |A_3| = |A_5| = |A_7| = a$, *and*

2. $|A_2| = |A_4| = |A_6| = b$.

The main fact is

Theorem 2.2.1. *Maximum m/M is achieved when K is a triangle and l_0, l_1, l_2 form a balanced partition on K.*

Proof.

Before we start, let us introduce several notions. As always, l_0, l_1, l_2 partition K into seven regions $A_1, A_2, \cdots, A_7$, m is the mininum area, M is the maximum area. We say that A_i is *relaxed* if $m < |A_i| < M$.

During the proof, we may either perturb partition lines l_0, l_1, l_2 to l'_0, l'_1, l'_2 or perturb convex body K to K'. After the perturbation, l'_0, l'_1, l'_2 partition K (or l_0, l_1, l_2 partition K') into seven regions $A'_1, A'_2, \cdots, A'_7$, m' is the mininum area, M' is the maximum area. A'_i is *relaxed* if $m' < |A'_i| < M'$. Also, let us denote $\Delta|A_i| = |A'_i| - |A_i|$.

There are special kinds of perturbations applying to K and one of its region A_i:

1. only the boundary of K inside region A_i is changed,

2. K' is still a convex body,

3. $|A'_i| < |A_i|$ (or $|A'_i| > |A_i|$),

4. $A_j = A'_j$ for $j \neq i$.

If $|A'_i| < |A_i|$, we call it a *squeeze*. If $|A'_i| > |A_i|$, we call it an *inflate*. If A_i can not be squeezed, it is *minimized*. Similarly, if A_i can not be inflated, it is *maximized*.

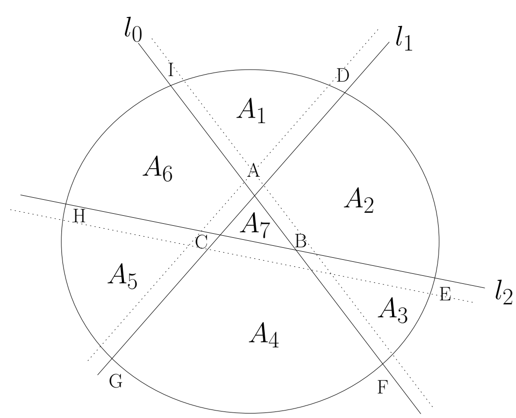

Figure 2.2: If $|A_7| < \min_{1 \le i \le 6}\{|A_i|\}$, we can perturb l_0, l_1, l_2 to improve the ratio.

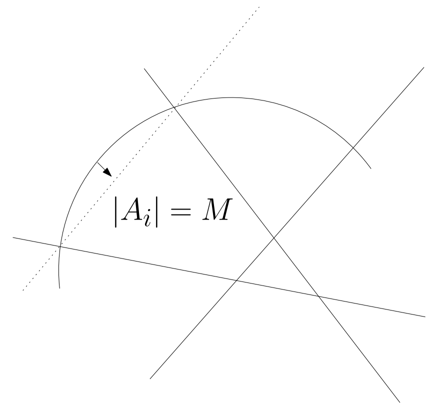

Figure 2.3: If $|A_i| = M$ and A_i is not minimized, squeeze A_i.

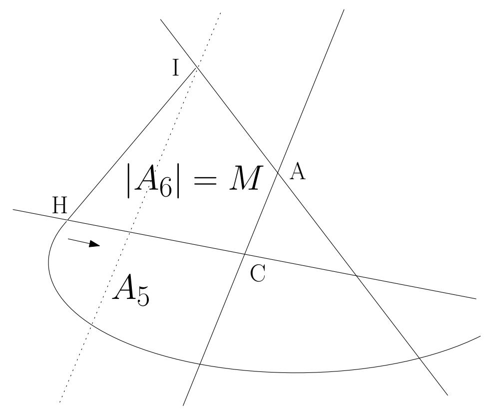

Figure 2.4: The case $|A_6| = M$ and $|A_5| > m$.

We will prove the theorem by establishing a sequence of numbered statements. Together they will prove the fact that unless K is a triangle and the three lines form a balanced partition, either K may be changed (making it closer to a triangle and increasing m/M) or one or more of the lines may be moved (making the partition more balanced and increasing m/M).

1. If m/M is maximum, then $\min_{1\leq i\leq 6}\{|A_i|\} \leq |A_7| \leq \max_{1\leq i\leq 6}\{|A_i|\}$.

   Suppose $|A_7| < \min_{1\leq i\leq 6}\{|A_i|\}$. Translate l_0 (resp. l_1, l_2) towards D (resp. H, F) by a small distance ϵ (as shown in Figure 2.2). It is clear that $|A_7|$ increases by $(\overline{AB} + \overline{BC} + \overline{CA})\epsilon + o(\epsilon)$ and $|A_1|, |A_3|, |A_5|$ decrease by a small amount. Our

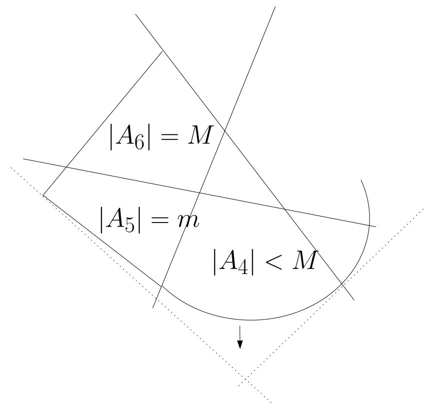

Figure 2.5: The case $|A_6| = M$, $|A_5| = m$, $|A_4| < M$ and A_4 is not minimized.

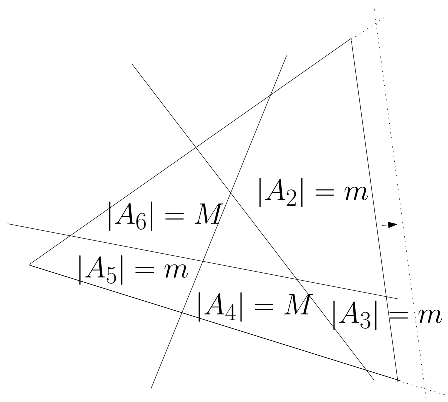

Figure 2.6: One of A_1, A_3 is maximized and minimized, $|A_2| = m$.

only concern is that $|A_2|, |A_4|, |A_6|$ may increase. Let us consider A_2. Because

$$\Delta|A_2| = (\overline{AD} + \overline{BE} - \overline{AB})\epsilon + o(\epsilon),$$

and

$$\frac{\overline{AD}}{\overline{AC}} < \frac{M}{m}, \frac{\overline{BE}}{\overline{BC}} < \frac{M}{m},$$

we have

$$\frac{m'}{M'} < \frac{m}{M},$$

contradicting the assumption that m/M was maximal.

2. If m/M is maximum, then $|A_1| = |A_3| = |A_5|$, $|A_2| = |A_4| = |A_6|$ and K is a triangle.

(a) Pick some $A_i, 1 \leq i \leq 6$, such that $|A_i| = M$.

(b) If A_i is not minimized, squeeze A_i (see Figure 2.3) and go to (a).

(c) Without loss of generality, assume $i = 6$, we have A_6 is minimized and $|A_6| = M$. Assume line HI is parallel with line AC or they intersect at the "upper" side, then $|A_5| = m$, or we could decrease $|A_6|$ and $|A_5|$ simultaneously (see Figure 2.4) and go to (a). Also, A_5 must be maximized, or we could first increase $|A_5|$ to $m + \epsilon_1$, then use the previous argument.

(d) With the similar argument as (c), we can conclude that one of the following two cases must happen.

   i. Either $|A_1| = m$ and A_1 is maximized, or

   ii. $|A_1| > m$, A_1 is maximized and minimized, $|A_2| = m$.

   Otherwise, we can decrease $|A_6|$ without increasing m and then go to (a).

(e) A_4 must be minimized, otherwise

   i. $|A_4| = M$, we can decrease $|A_4|$ and increase $|A_5|$, then go to (c),

   ii. $|A_4| < M$, we can keep or increase $|A_4|$ and increase $|A_5|$ at the same time (see Figure 2.5), then go to (c).

(f) With the similar argument as (e), we can conclude that one of the following two cases must happen

   i. $|A_4| = M$ and A_4 is minimized.

   ii. $|A_4| < M$, A_4 is maximized and minimized, $|A_3| = M$ and A_3 is minimized, $|A_2| = m$ and A_2 is maximized.

   Furthermore, the second case is impossible, because if we consider it together with two cases in (d), we can only reach two impossible configurations. Therefore, $|A_4| = M$.

(g) Finally, we consider A_3. Similar to (d), either $|A_3| = m$ and A_3 is maximized, or $|A_3| > m$, A_3 is maximized and minimized, $|A_2| = m$. Combining with (d), we have three cases:

i. $|A_1| = |A_3| = m$ and both are maximized. In this case, it is clear that A_2 must be minimized and $|A_2| = M$, which is exactly what we want to prove.

ii. Exactly one of A_1, A_3 has area m and is maximized. The other is maximized and minimized, $|A_2| = m$. As shown in Figure 2.6, we can increase two minimum areas and then go to (d).

iii. Both A_1, A_3 are maximized and minimized, $|A_2| = m$. We can increase A_2 in this case and then go to (d).

3. If m/M is maximum, then $|A_1| = |A_3| = |A_5| = |A_7| = m$, $|A_2| = |A_4| = |A_6| = M$ and K is a triangle.

(a) It is easy to check that $|A_1| = |A_3| = |A_5| = M$ and $|A_2| = |A_4| = |A_6| = m$ is an impossible configuration.

(b) If $|A_1| = |A_3| = |A_5| = m$, $|A_2| = |A_4| = |A_6| = M$ and $|A_7| > m$, we can improve the ratio by perturbing l_0, l_1, l_2.

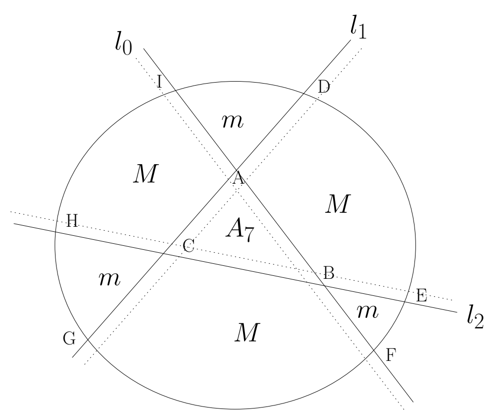

Figure 2.7: $|A_1| = |A_3| = |A_5| = m$, $|A_2| = |A_4| = |A_6| = M$ and $|A_7| > m$.

As shown in Figure 2.7, let us translate l_0 (resp. l_1, l_2) towards C (resp. B, A) by a small distance $\frac{\epsilon M}{AB}$ (resp. $\frac{\epsilon M}{AC}$, $\frac{\epsilon M}{BC}$). We have $|A_7'| < |A_7|$ and $|A_i'| < |A_i|(1+\epsilon)$ for $i = 2, 4, 6$. Therefore, to prove that $m'/M' > m/M$, all we need to show is $|A_i'| \geq |A_i|(1+\epsilon)$ for $i = 1, 3, 5$. We prove that by contradiction.

Suppose $|A_1'| < |A_1|(1+\epsilon)$. Then we have

$$\frac{|A_1'| - |A_1|}{|A_1|} = \frac{M\epsilon(\frac{\overline{AD}}{\overline{AC}} + \frac{\overline{AI}}{\overline{AB}})}{m} < \epsilon,$$

i.e.,

$$\frac{\overline{AD}}{\overline{AC}} + \frac{\overline{AI}}{\overline{AB}} < \frac{m}{M}.$$

It is easy to see that $\frac{\overline{AD}}{\overline{AC}} < \frac{m}{M}$ implies H is closer to l_1 than I, so $\frac{\overline{CH}}{\overline{CB}} < \frac{\overline{AI}}{\overline{AB}} < \frac{m}{M}$. This implies that G is closer to l_2 than F. Similarly, we have $\frac{\overline{BE}}{\overline{BC}} < \frac{\overline{AD}}{\overline{AC}} < \frac{m}{M}$, which implies F is closer to l_2 than G. Thus we have a contradiction.

2.3 Uniqueness proof

Theorem 2.3.1. *Given a convex body $K \subset \mathbb{R}^2$ with $|K| = 1$ and a unit vector $v \in \mathbb{R}^2$, there are exactly two triples of lines that form a balanced partition of K, where one of the three lines has normal vector v.*

Proof.

Without loss of generality we may take the given normal vector to be $v = (1, 0)$, so one of the three lines will be a vertical line which we denote by l_0. It will have equation $x = t$ and we write $l_0(t)$ to describe its position. Futhermore, we coordinatize $\mathbb{R}^2$ so that $l_0(0)$ bisects K, i.e., K has area $1/2$ on both sides of $l_0(0)$.

Our proof goes in two steps:

1. First, we construct a unique "candidate" for a balanced partition at t, which is a triple of lines $l_0(t), l_1, l_2$, such that any other triple of lines $l_0(t), l_1', l_2'$, where $(l_1', l_2') \neq (l_1, l_2)$, can not be a balanced partition of K.

2. Second, we prove that exactly two "candidates" are balanced partitions for all possible t.

Step 1. Construct a unique candidate at t.

We will focus on the case $t \geq 0$. For $t < 0$, a similar construction and argument will apply.

Given some $t \geq 0$, we write K_0^- for the smaller (right) part of K cut off by $l_0(t)$, K_0^+ to be the bigger(left) part of K cut off by $l_0(t)$ and $\lambda = |K_0^-| \leq 1/2$. It is clear that $l_0(t), l_1, l_2$ can form a balanced partition of K **only if** there exists non-negative numbers a and b, which satisfy

$$\begin{cases} 2a & + & b & = & \lambda, \\ 2a & + & 2b & = & 1-\lambda, \end{cases}$$

where $a = |A_1| = |A_3| = |A_5| = |A_7|$ and $b = |A_2| = |A_4| = |A_6|$ (see Figure 2.1).

Clearly,

$$\begin{cases} a & = & \frac{3\lambda-1}{2}, \\ b & = & 1-2\lambda. \end{cases} \tag{2.1}$$

By Equation 2.1, both a and b are non-negative if and only if $\lambda \in [1/3, 1/2]$. Notice that λ decreases from $1/2$ to 0 as t goes from 0 to $+\infty$, so a and b are determined uniquely by t and they are both non-negative if and only if $t \in [0, t_{1/3}]$, $t_{1/3}$ denotes the t value where corresponding λ is $1/3$.

Now, for every $t \in [0, t_{1/3}]$, we define l_1 and l_2. By the separated ham-sandwich theorem (see [1], [5]), the following equations uniquely define l_1 and l_2 (as shown in Figure 2.8).

$$\begin{cases} |l_1^+ \cap K_0^-| = a, & |l_1^+ \cap K_0^+| = a+b, \\ |l_2^- \cap K_0^-| = a, & |l_2^- \cap K_0^+| = a+b, \end{cases} \tag{2.2}$$

where $l_i^+(l_i^-)$ is the halfsapce above(below) $l_i, i = 1, 2$, a and b are non-negative numbers determined by t and the convex body K.

On the other hand, it is clear that Equation 2.2 is a necessary (but not sufficient) condition for $l_0(t), l_1, l_2$ to be a balanced partition of K. Therefore, for every $t \in [0, t_{1/3}]$, a unique "candidate" for a balanced partition at t, which is a triple of lines $l_0(t), l_1, l_2$, can be constructed (as in Figure 2.8). Similar results apply for $t \in [t_{-1/3}, 0]$, where $t_{-1/3}$ denotes the t for which $l_0(t)$ cuts $1/3$ off K on its left and $2/3$ off K on its right.

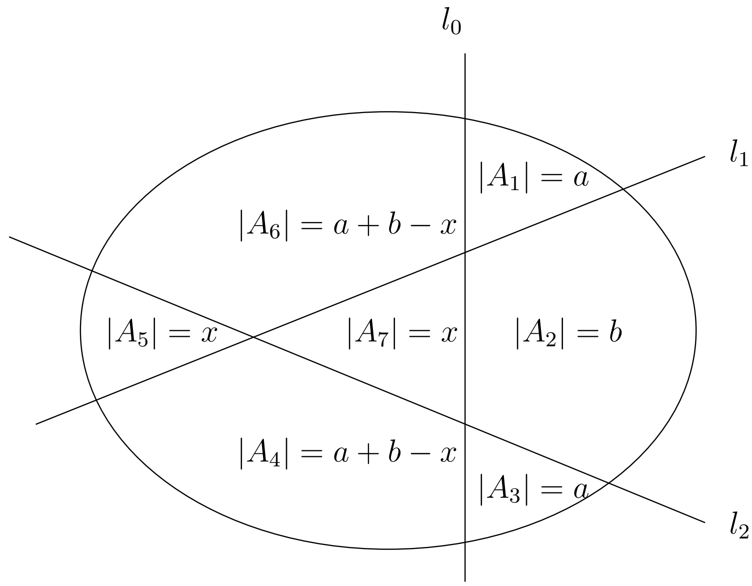

Figure 2.8: $l_0(t), l_1, l_2$ is the "candidate" balanced partition at t.

Step 2. Among all $t \in [t_{-1/3}, t_{1/3}]$, exactly two "candidates" are balanced partitions.

Suppose the unique "candidate" at t will partition K into seven regions $A_1, A_2, \ldots, A_7$ (as in Figure 2.8), let us define the function

$$f(t) = a - |A_5|, t \in [t_{-1/3}, t_{1/3}].$$

Clearly, $f(t) = 0$ is both necessary and suffiicent for the "candidate" at t to be a balanced partition. Therefore, to prove there are exactly two "candidates" which are balanced paritions of K, we only need to prove that there are exactly two roots of $f(t)$ among $t \in [t_{-1/3}, t_{1/3}]$.

The following properties of $f(t)$ are enough to show that $f(t)$ has exactly two roots (as illustrated in Figure 2.9):

1. $f(t)$ is continuous on $[t_{-1/3}, t_{1/3}]$,

2. $f(0) > 0$, $f(t_{-1/3}) < 0$, $f(t_{1/3}) < 0$,

3. $f(t)$ is strictly increasing on $[t_{-1/3}, 0]$ and strictly decreasing on $[0, t_{1/3}]$.

Suppose t and t' are very close. $l_0(t), l_1, l_2$ form the candiate at t and $l_0(t'), l_1', l_2'$ form the candiate at t'. Clearly, $l_0(t)$ will be very close (in distance) to $l_0(t')$, λ will

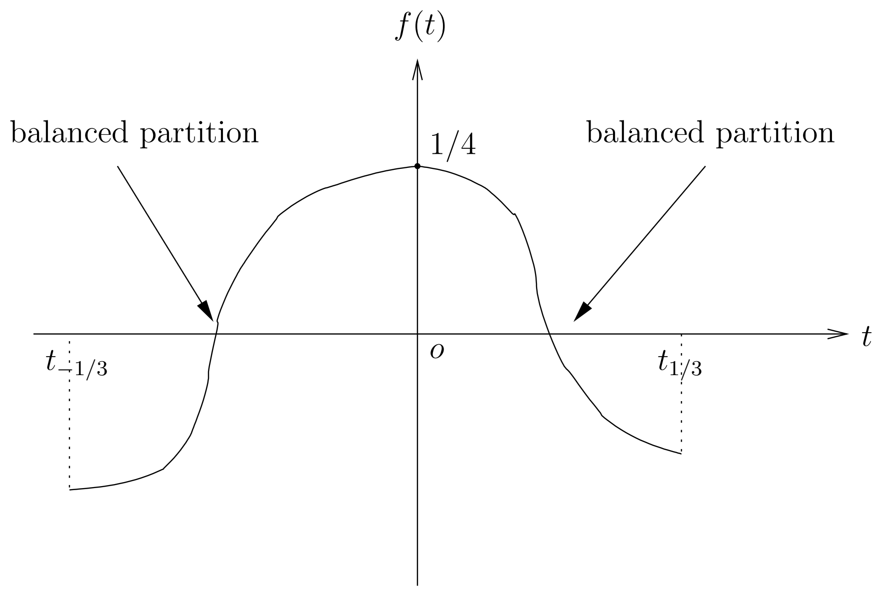

Figure 2.9: Two roots of $f(t)$ corresponding to two "candidates" which form balanced partitions of K.

be very close to λ' and a (b) will be very close to a' (b'). As l_1 and l_2 (l'_1 and l'_2) are seperated ham-sandwich cuts determined by l_0, a and b (l'_0, a' and b'), l_1 will be very close (in distance) to l'_1 and l_2 will be very close (in distance) to l'_2. Furthermore, x will be very close to x' because x is determined by K, l_1 and l_2 and x' is determined by K, l'_1 and l'_2. Therefore, $f = a - x$ will be very close to $f' = a' - x'$, which implies Property 1.

To prove property 2, we consider two boundary cases, $t = 0$ and $t = t_{1/3}$. The case that $t = t_{-1/3}$ is similar to $t = t_{1/3}$.

When $t = 0$, $a = 1/4$ and $b = 0$, $f(0) = 1/4 - 0 = 1/4$, as illustrated in Figure 2.10.

When $t = 1/3$, $a = 0$ and $b = 1/3$, $f(t_{1/3}) = 0 - |A_5| < 0$, as illustrated in Figure 2.11. One thing worth notice is why $|A_5| > 0$. Is it possible that l_1 and l_2 do not intersect inside K so that A_5 vanishes and $|A_5| = 0$? The answer is no. Assume l_1 and l_2 do not intersect inside K, we have $|K_0^-| > |A_4| + |A_6| = 1/3 + 1/3 = 2/3$, which contradicts the fact that $|K_0^-| = 2/3$ when $t = 1/3$.

We will prove property 3 for $t \in [0, t_{1/3}]$; the proof for $t \in [t_{-1/3}, 0]$ is analogous.

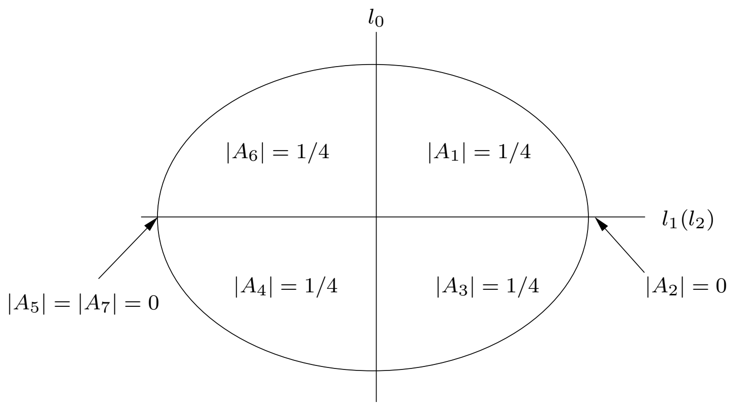

Figure 2.10: the "candidate" balanced partition when $t = 0$, l_1 and l_2 overlap.

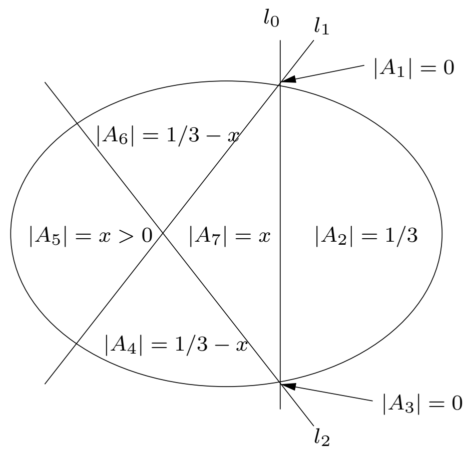

Figure 2.11: the "candidate" balanced partition when $t = 1/3$.

Given $0 \leq t < t' \leq t_{1/3}$, let $l_0(t), l_1, l_2$ be the "candidate" at t, with the corresponding parameters λ, a and b; let $l_0(t'), l_1', l_2'$ be the "candidate" at t', with the corresponding parameters λ', a' and b'. Let us say that $l_0(t), l_1, l_2$ partition K into seven regions $A_1, A_2, \cdots, A_7$ and $l_0(t'), l_1', l_2'$ partition K into seven regions $A_1', A_2', \cdots, A_7'$.

It is clear that $\lambda' < \lambda$. So Let us assume that $\lambda' = \lambda - \epsilon$ with some $\epsilon > 0$. By equation 2.1, we have

$$\begin{cases} a' &= a - \frac{3\epsilon}{2}, \\ b' &= b + 2\epsilon. \end{cases}$$

Notice that $|A_1| - |A_1'| = a - a' = 3\epsilon/2$, which is the result of moving $l_0(t)$ to $l_0(t')$ and moving l_1 to l_1', as shown in Figure 2.12. Because moving $l_0(t)$ to $l_0(t')$ decreases

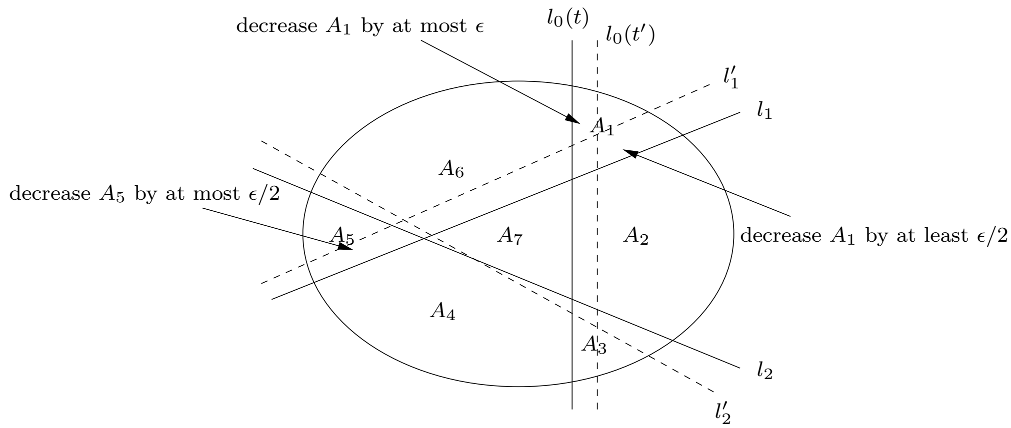

Figure 2.12: Moving l_1 to l'_1 can not decrease the area of A_5 by more than $\epsilon/2$.

the area of A_1 by at most ϵ, we have that moving l_1 to l'_1 decreases the area of A_1 by at least $\epsilon/2$. On the other hand, moving l_1 to l'_1 decreases the area of $A_1 \cup A_5 \cup A_6$ by ϵ (from λ to λ'). So we can conclude that moving l_1 to l'_1 can not decrease the area of A_5 by more than $\epsilon/2$.

Similarly, we can show that moving l_2 to l'_2 can not decreases the area of A_5 by more than $\epsilon/2$, too. Therefore, $|A_5| \leq |A'_5| + \epsilon + o(\epsilon)$.

Combine the previous observations together, we have

$$a - |A_5| \geq a' + \frac{3\epsilon}{2} - |A'_5| - \epsilon - o(\epsilon) > a' - |A'_5|,$$

i.e., $f(t) > f(t')$, as claimed.

2.4 Optimal partition on triangles

By Lemma 2.2.1, we know that when m/M is maximum, K must be a triangle, say ΔPQR. Since ΔPQR may be made equilateral by an affine transformation and an affine transformation preserves the ratio of areas, we may assume ΔPQR is an equilateral triangle centered at point O. Let $T = \{l_0, l_1, l_2\}$ partition ΔPQR into seven regions and let ΔABC be the center region (see Figure 2.13).

The following lemma shows that ΔABC is an equilateral triangle centered at point O, if T is a balanced partition of ΔPQR.

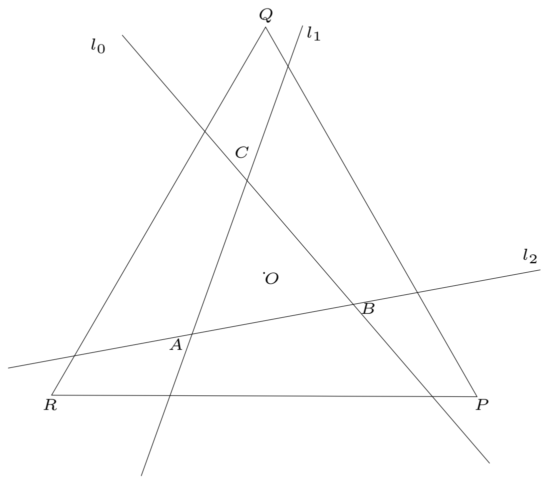

Figure 2.13: l_0, l_1, l_2 partition ΔPQR into seven regions.

Lemma 2.4.1. *If $T = \{l_0, l_1, l_2\}$ forms a balanced partition of ΔPQR, then ΔABC is an equilateral triangle centered at point O.*

Proof.

Given any unit vector $v = (\sin\theta, \cos\theta)$, let $v_1 = (\sin(\theta - \frac{\pi}{3}), \cos(\theta - \frac{\pi}{3}))$ and $v_2 = (\sin(\theta - \frac{2\pi}{3}), \cos(\theta - \frac{2\pi}{3}))$. l_0 (resp., l_1, l_2) is initialized to be with normal vector v (resp., v_1, v_2) and passing through the point O, the center of ΔPQR (see Figure 2.14).

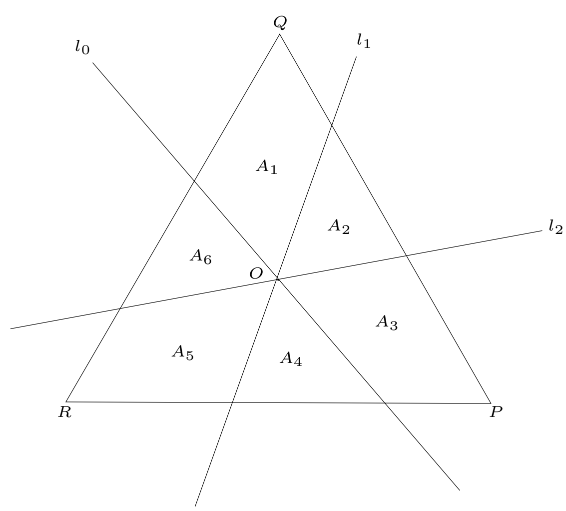

Figure 2.14: l_0 (resp., l_1, l_2) is initialized to have normal vector v (resp., v_1, v_2) and pass through the point O.

It is clear that $|A_1| = |A_3| = |A_5|$, $|A_2| = |A_4| = |A_6|$ and $|A_7| = 0$.

Now let us translate l_0 towards Q, l_1 towards R and l_2 towards P at the same linear speed until the area of the center region equals one of the six outer regions. We denote this triple of lines by $T^+(v)$.

We claim that $T^+(v)$ has the following properties:

1. the center region is an equilateral triangle centered at O.

2. one line of $T^+(v)$ has normal vector v.

3. $T^+(v)$ forms a balanced partition of ΔPQR.

The first two properties are trivially true. We will prove the third property here.

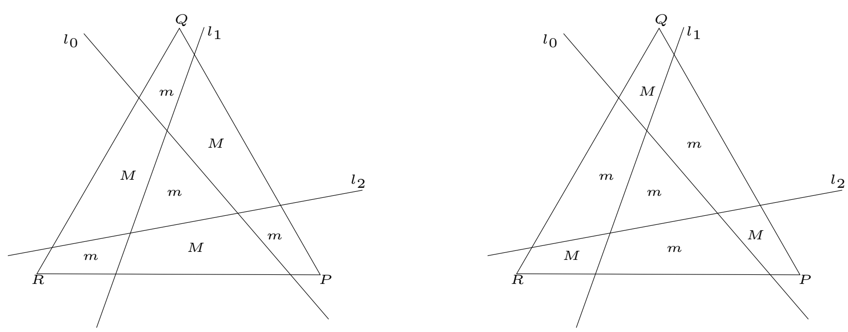

Figure 2.15: Two cases for $T^+(v)$.

It is clear that for $T^+(v)$, either $|A_1| = |A_3| = |A_5| = |A_7| = m$, $|A_2| = |A_4| = |A_6| = M$ or $|A_1| = |A_3| = |A_5| = M$, $|A_2| = |A_4| = |A_6| = |A_7| = m$, as shown in Figure 2.15. $T^+(v)$ is a balanced partition of ΔPQR in the first case. All we need to show is that the second case is impossible.

Suppose that QP is parallel to or converges towards CB as in Figure 2.16. Let $E'DP'$ be parallel to CB, we have

$$\frac{areaABGF}{areaBDPG} > \frac{areaABG}{areaBDP'G} = \frac{areaACB}{areaBCE'D} \geq \frac{areaACB}{areaBCED},$$

i.e., $\frac{|A_4|}{|A_3|} > \frac{|A_7|}{|A_2|}$.

If $|A_1| = |A_3| = |A_5| = M$, $|A_2| = |A_4| = |A_6| = |A_7| = m$, we have $\frac{m}{M} > \frac{m}{m}$, which is a contradiction.

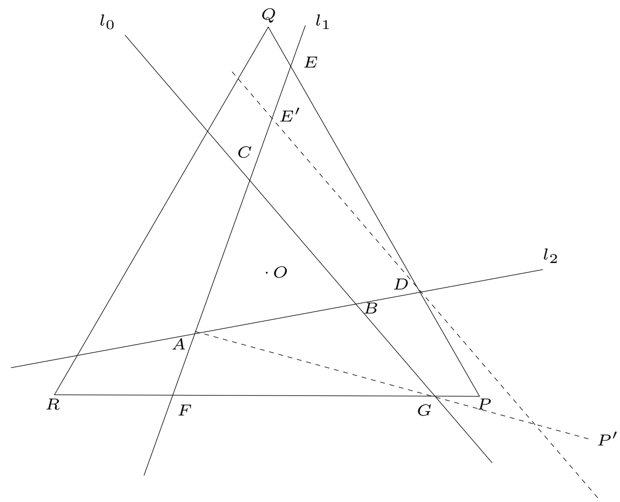

Figure 2.16: The second case is an impssible configuration.

Similarly, we may define $T^-(v)$, by sliding l_0, l_1 and l_2 towards the other direction. By the same argument, we can show that $T^-(v)$ has the same properties as $T^+(v)$.

By Theorem 2.3.1, $T^+(v)$ and $T^-(v)$ are the only two balanced partitions of ΔPQR, containing one line with normal vector v. Thus, we may conclude that every balanced partition of ΔPQR has the property that the center region is an equilateral triangle centered at O.

By Lemma 2.2.1 and Lemma 2.4.1, we have the following lemma.

Lemma 2.4.2. *If $T = \{l_0, l_1, l_2\}$ achieves maximum m/M on ΔPQR, then ΔABC is an equilateral triangle centered at point O.*

Lemma 2.4.3. *If $T = \{l_0, l_1, l_2\}$ achieves maximum m/M on ΔPQR, then AB (resp., BC, CA) is parallel to RP (resp., PQ, QR).*

Proof.

First of all, P , Q and R must belong to regions of area m. Otherwise, there is some region of area M, whose area may be reduced without violating the convexity (see Figure 2.17). In this way, we get some convex body K in which m/M is the same as (or better than) maximum m/M and K is not a triangle. This contradicts the result of Lemma 2.2.1.

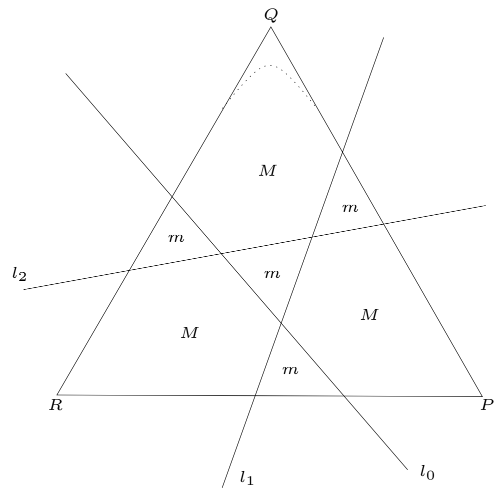

Figure 2.17: If P, Q and R do not belong to regions of area m, some regions of area M may be reduced without violating the convexity.

Now let us assume AB (resp., BC, CA) is not parallel to RP (resp., PQ, QR). By Lemma 4, ΔABC is an equilateral triangle centered at point O. As shown in Figure 2.18, let the perpendicular bisector of AB (resp., BC, CA) meet RP (resp., PQ, QR) at C' (resp., A', B'). Let $\Delta P'Q'R'$ satisfy the following:

1. $\Delta P'Q'R'$ is an equilateral triangle centered at O.

2. $P'Q'$ (resp. $Q'R'$, $R'P'$) passing through A' (resp., B', C').

3. $\angle Q'A'Q = \angle R'B'R = \angle P'C'P = \epsilon$, where ϵ is a very small angle.

l_0, l_1, l_2 partitions ΔPQR into seven regions A_1 (CEQH), A_2 (CBDE), $\cdots$, A_7 (ABC). Correspondingly, l_0, l_1, l_2 partitions $\Delta P'Q'R'$ into A'_1 ($CE'QH'$), A'_2 ($CBD'E'$), $\cdots$, A'_7 (ABC). Clearly, we have

1. $|A_1| = |A_3| = |A_5| = |A_7| = m$, $|A_2| = |A_4| = |A_6| = M$,

2. $|A'_1| = |A'_3| = |A'_5|$, $|A'_2| = |A'_4| = |A'_6|$,

3. $|A'_7| = |A_7|$.

   Furthermore, if $P'Q'$ is closer to being parallel to BC, we have

4. $|A'_2| = |A'_4| = |A'_6| < M$,

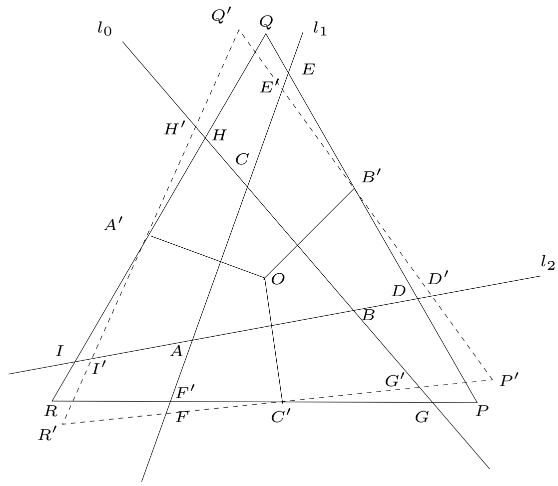

Figure 2.18: AB (resp., BC, CA) is not parallel to RP (resp., PQ, QR)

5. $|A_1'| = |A_3'| = |A_5'| > m.$,

Property 4 is obviously true, because $|\Delta A'E'E| > |\Delta A'D'D|$ by elementary geometry. Now we prove property 5 by showing $|A_1'| > |A_1|$. Since

$$|A_1'| - |A_1| = (|\Delta B'Q'A'| - |\Delta B'QA'|) + (|\Delta A'E'E| - |\Delta B'H'H|),$$

it is enought to show that $|\Delta B'Q'A'| > |\Delta B'QA'|$ and $|\Delta A'E'E| > |\Delta B'H'H|$. Notice that $\angle B'Q'A' = \angle\Delta B'QA' = \pi/3$, so Q and Q' lie on a circle which has $A'B'$ as a chord. It is easy to see that the distance from Q' to $A'B'$ is greater than the distance from Q to $A'B'$, thereofore, $|\Delta B'Q'A'| > |\Delta B'QA'|$. Finally, $|\Delta A'E'E| > |\Delta B'H'H|$ is true because $|\Delta B'H'H| = |\Delta A'D'D|$ and $|\Delta A'E'E| > |\Delta A'D'D|$.

Thus, with l_0, l_1, l_2, $\Delta P'Q'R'$ may achieve a ratio greater than m/M, which contradicts the assumption that m/M is maximized.

By Lemmas 2.4.2 and 2.4.3, we have the following theorem.

Theorem 2.4.1. *Given a convex body $K \subset \mathbb{R}^2$ and a triple of lines $T = l_0, l_1, l_2$ which partition K into seven regions $A_1, A_2, \cdots, A_7$, we have*

$$\frac{m}{M} \leq \frac{2}{1+2\sqrt{2}} \approx 0.522,$$

where $m = min\{|A_i|\}$ and $M = max\{|A_i|\}$. Equality holds only when K is a triangle,

l_0, l_1, l_2 are parallel to three sides of K and form a balanced partition of K as in Figure 2.19. (Again, the ratio is preserved under affine transformation.)

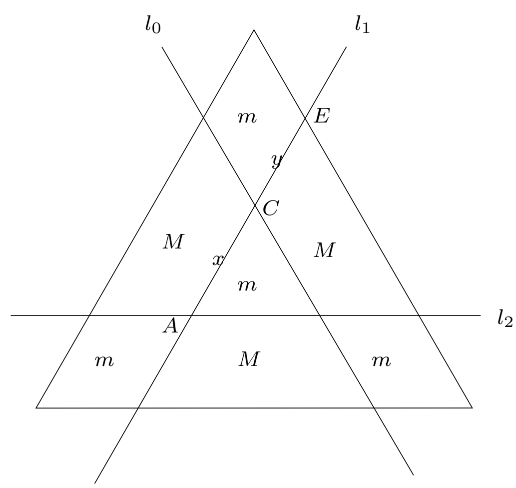

Figure 2.19: The configuration of K and l_0, l_1, l_2 when optimal ratio is achieved.

Proof.

We have shown that if m/M is maximum, then K is a triangle, l_0, l_1, l_2 are parallel to three sides of K and form a balanced partition of K.

Now let us compute the maximum m/M. Let $x = |AC|$ and $y = |CE|$, it is clear that $x^2 = 2y^2$. Thus we have

$$\frac{m}{M} = \frac{x^2}{(y+x)^2 - x^2} = \frac{2}{1 + 2\sqrt{2}} \approx 0.522.$$

Chapter 3

Zigzag - Traversing Points in the Plane

3.1 Definition of the zigzag problem

Definition Let P be a set of points in general position such that no three points of P are colinear. A *(simple) zigzag* z of P is a (simple) polygonal chain which passes through every point $v \in P$.

Notice that A *(simple) zigzag* z of P may "turn" at points in P or at points not in P. We denote the length of a (simple) zigzag z as $|z|$, which is the number of line segments in z.

Definition Let Z^P denote the length of the shortest zigzag of P, i.e., $Z^P = \min\{|z| : z \text{ is a zigzag of } P\}$. Similarly, $Z_s^P = \min\{|z| : z \text{ is a simple zigzag of } P\}$.

Furthermore, $Z(n) = \max\{Z^P : P \text{ is a set of } n \text{ points in general position}\}$. Similarly, $Z_s(n) = \max\{Z_s^P : P \text{ is a set of } n \text{ points in general position}\}$.

It is easy to show that $\lceil \frac{n}{2} \rceil \leq Z(n) \leq Z_s(n) \leq n-1$. We will improve these trivial bounds in the following sections.

3.2 Zigzag with self-intersection

Theorem 3.2.1.

$$Z(n) = \lceil \frac{n}{2} \rceil + O\left(\frac{n}{\log n}\right).$$

Proof. Let us start with a simple observation that k points in convex position can be traversed by a zigzag of size $\lceil (k+1)/2 \rceil$. Besides, if P_1 and P_2 forms a partition of point set P, by connecting an endpoint of the shortest zigzag of P_1 to an endpoint of the shortest zigzag of P_2 with one more extra line segments, we have a zigzag of P with length $Z_{P_1} + Z_{P_2} + 1$. So

$$Z_P \leq Z_{P_1} + Z_{P_2} + 1.$$

Therefore, if any set of n points in the plane in general position has a subset of $k(n)$ points in convex position, then we have

$$Z(n) \leq k(n)/2 + Z(n - k(n)) + 2. \tag{3.1}$$

Let $ES(k)$ denote the least integer such that among any $ES(k)$ points in general position in the plane there are always k in convex position. By Erdös-Szekeres Theorem, $ES(k)$ exists and

$$ES(k) \leq \binom{2k-4}{k-2} + 1.$$

Thus, we have

$$k(n) = \Omega(\log n). \tag{3.2}$$

Combining 3.1 and 3.2, we may conclude that

$$Z(n) = \lceil \frac{n}{2} \rceil + O\left(\frac{n}{\log n}\right).$$

□

3.3 Zigzag without self-intersection

Theorem 3.3.1.

$$Z_s(n) \leq n - \lfloor \frac{n-2}{8} \rfloor.$$

Proof.

Consider P, a set of n points in general position in the plane. For simplicity, we assume $n = 8k+2$. First of all, we pick a direction d which is not parallel or perpendicular to any line passing through a pair of points in P. Now points in P can be sorted by direction d, that is, $P = \{v_0, v_1, \cdots, v_{8k+1}\}$, such that v_i is "above" v_j in direction d for every $0 \leq i < j < n$. Let $l(v_i)$ denote the line passing through v_i and perpendicular to d for $0 \leq i \leq 8k$. Let $T(v_i) = R$ if line segments $\overline{v_{i-1}v_i}$ and $\overline{v_iv_{i+1}}$ turn right and $T(v_i) = L$ otherwise, for $0 < i < 8k+1$.

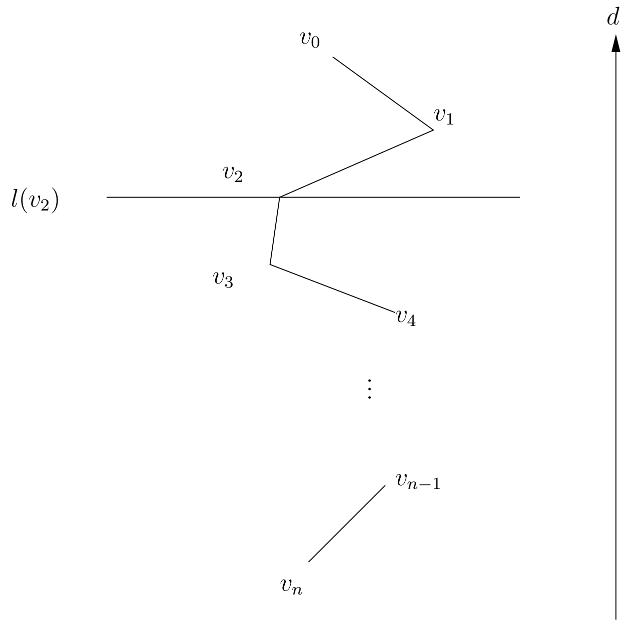

Figure 3.1: Sort points by direction d.

Obviously, there is a trivial zigzag $v_0 v_1 \cdots v_{8k}$ with $n - 2 = 8k$ turns. Our goal is to improve the trivial zigzag by removing k turns. The idea is to "slice" the plane into k consecutive horizontal slabs and revise the trivial zigzag in each slab so that at least one turn is saved. Formally, the ith slab contains the area between $l(v_{8i+1})$ and $l(v_{8i+9})$, 9 points $\{v_{8i+1}, v_{8i+2}, \cdots, v_{8i+9}\}$ and a part of the trivial "zigzag" with 8 turns (the turn at v_{8i+9} belongs to the $(i+1)$th slab), where $0 \le i < k$. For each slab, we construct a new zigzag which still starts from v_{8i+1}, ends in v_{8i+9} and has at least one less turn than the trivial zigzag. Furthermore, this zigzag should be inside the area of the ith slab to promise that zigzags of different slabs will not intersect.

Proposition 3.3.2. *Let d be a direction and $u_0, u_1, \cdots, u_9$ be* 10 *points in general poisition, so that u_i is "above" u_j in direction d for every $0 \le i < j \le 9$. There always exists a zigzag z that starts from u_1, ends in u_9 and totally falls inside the area between $l(u_1)$ and $l(u_9)$, so that the zigzag $z' = \overline{u_0 u_1} + z$ contains at most 7 turns. We call such z a* nice zigzag.

Here, $u_1, u_2, \cdots, u_9$ correspond to $\{v_{8i+1}, v_{8i+2}, \cdots, v_{8i+9}\}$ in the ith slab, u_0 refers to the second to last (the one before v_{8i+1}) point in P visited by the zigzag constructed

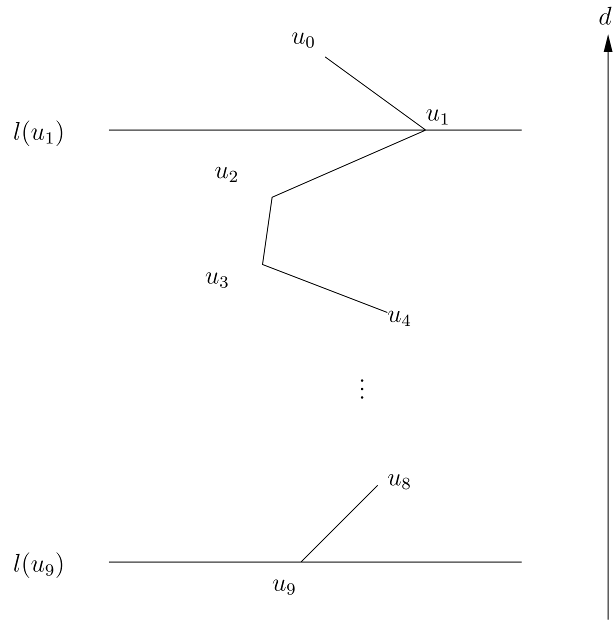

Figure 3.2: Trivial zigzag in one slab.

in the previous slab(if $i = 0$, then $u_0 = v_0$). If we simply use the trivial zigzag, it will cost 8 turns.

We will prove the previous lemma by a sequence of lemmas, here is the first one.

Lemma 3.3.1. *If* $T(u_i) = T(u_{i+1})$ *for some* i, $1 \leq i < 9$, *then a nice zigzag exists.*

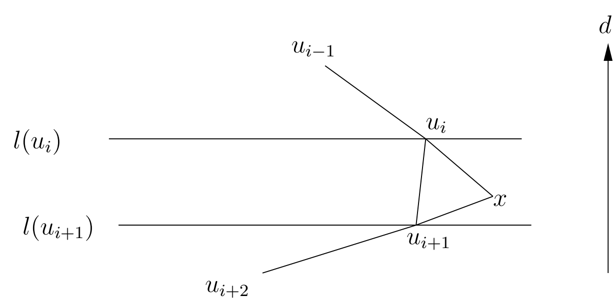

Figure 3.3: Two consecutive right turns.

Proof. Suppose $T(u_i) = T(u_{i+1}) = R$ for some i, $1 \leq i < 9$. As shown in Figure 3.3, line $u_{i-1}u_i$ must intersect line $u_{i+1}u_{i+2}$ at some point x between $l(u_i)$ and $l(u_{i+1})$. Take the trivial zigzag, replace line segment u_iu_{i+1} by u_ixu_{i+1} and we have a nice zigzag. $\square$

Without loss of generality, from now on, we may assume $T(u_i) = R$ when i is odd and $T(u_i) = L$ when i is even for $1 \leq i < 9$.

Lemma 3.3.2. *For any i such that $2 \leq i \leq 6$, let a_i denote the intersection of line $u_{i-1}u_i$ with $l(u_9)$ and b_i denote the intersection of line u_iu_{i+1} with $l(u_9)$. If u_{i+2} does not lie inside $\triangle u_ia_ib_i$, then a nice zigzag exists.*

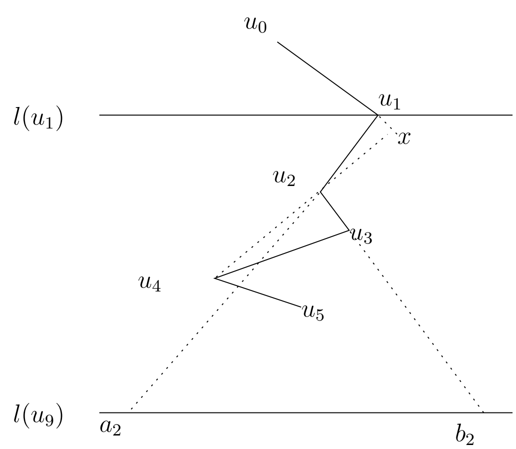

Figure 3.4: u_4 lies outside $\triangle u_2a_2b_2$.

Proof. We check the case of $i = 2$. For other i's, the proof is similar. As shown in Figure 3.4, if u_4 is outside $\triangle u_2a_2b_2$, it must be on the left side of line u_1u_2 because $T(u_3) = R$. Thus, line u_0u_1 intersects u_4u_2 at some point x between $l(u_1)$ and $l(u_2)$. It is easy to check that $u_1xu_4u_3u_5u_6u_7u_8u_9$ is a nice zigzag (u_3u_5 may not intersect other line segments as $T(u_4) = L$). □

Now we may assume each u_{i+2} lies inside $\triangle u_ia_ib_i$ for $2 \leq i \leq 6$, otherwise a nice zigzag already exists. As a matter of fact, by the same argument, we may assume each u_{i+4} lies inside $\triangle u_ia_ib_i$ as well, for $2 \leq i \leq 4$. Combining these two assumptions, we may conclude that, for $2 \leq i \leq 4$, ray $u_{i+4}u_{i+2}$ intersects either line segment $\overline{u_ia_i}$ or line segment $\overline{u_iu_{i+1}}$ at some point x. Besides, x is not in line segment $\overline{u_{i+4}u_{i+2}}$.

Lemma 3.3.3. *For $2 \leq i \leq 4$, if line $u_{i+4}u_{i+2}$ intersects line segment $\overline{u_ia_i}$ at some point x, then a nice zigzag exists.*

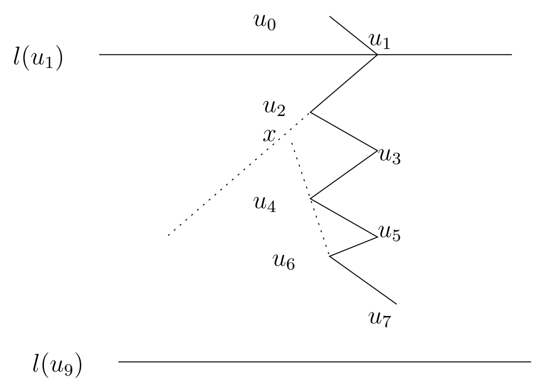

Figure 3.5: u_6u_4 intersects u_2a_2.

Proof. Again, a nice zigzag is constructed for $i = 2$, similar construction could be done for other i's. As shown in Figure 3.5, a nice zigzag can be constructed by passing through $u_1u_2xu_4u_6$. If u_5 and u_7 are on the same side of line u_3u_6, then $u_1xu_6u_3u_5u_7u_8u_9$ is a nice zigzag. Otherwise, $u_1xu_6u_5u_3u_7u_8u_9$ is a nice zigzag. □

Let us assume line $u_{i+4}u_{i+2}$ intersects line segment $\overline{u_iu_{i+1}}$ at some point x, for $2 \leq i \leq 4$. Consider y, the intersection of line u_iu_{i+2} with line $u_{i+4}u_{i+3}$. If y is not on line segment $\overline{u_{i+4}u_{i+3}}$, then y must be "above" u_{i+3} (as for direction d). The following lemma asserts that a nice zigzag could be found in this situation.

Lemma 3.3.4. *For $2 \leq i \leq 4$, if line $u_{i+4}u_{i+2}$ intersects line segment $\overline{u_iu_{i+1}}$ at some point x and line u_iu_{i+2} intersects line $u_{i+4}u_{i+3}$ at some point y which is "above" u_{i+3} (as for direction d), then a nice zigzag exists.*

Proof. For $i = 2$, as shown in Figure 3.6, $u_1u_3u_2yu_6u_7u_8u_9$ is a nice zigzag. The case $i = 3$ or $i = 4$ are similar. □

Now we deal with the last case.

Lemma 3.3.5. *If for every i, $2 \leq i \leq 4$, both line $u_{i+4}u_{i+2}$ intersects line segment $\overline{u_iu_{i+1}}$ and line u_iu_{i+2} intersects line segment $\overline{u_{i+4}u_{i+3}}$, then a nice zigzag exists.*

Proof. Line u_2u_4 must intersect line u_8u_6 at some point z inside $\triangle u_4u_5u_6$. If u_7 and u_9 lie at the same side of line u_8u_5, then $u_1u_3u_2zu_8u_5u_7u_9$ is a nice zigzag. Otherwise,

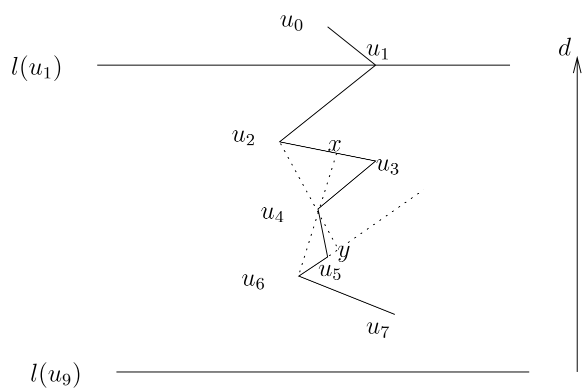

Figure 3.6: u_2u_4 intersects u_6u_5 at some y "above" u_5 .

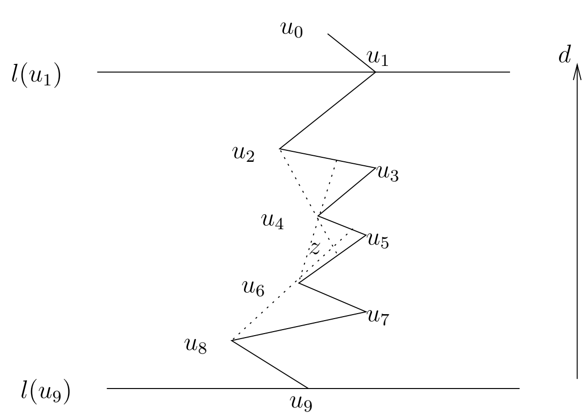

Figure 3.7: the last case.

$u_1u_3u_2zu_8u_7u_5u_9$ is a nice zigzag. In the latter case, line segment $\overline{u_5u_9}$ may not intersect other parts of the zigzag because $T(u_8) = L$. □

We believe that the $n - \lfloor \frac{n-2}{8} \rfloor$ upper bound for the non-intersection case can be improved. One possible approach is to apply the technique in our proof recursively such that more nice zigzags can be constructed.

Chapter 4

Optimally Balanced Forward Degree Sequence

4.1 Introduction

Forward degree seqences, arising from orderings of the vertices in a graph, carry a lot of vital information about the graph. In this chapter, we focus our work on two special classes of forward degree sequences, which we named balanced and strongly balanced. Our main result is to prove that any chordal graph has a strongly balanced forward degree sequence and any graph with all degrees at most 3 has a balanced forward degree sequence. Moreover, we show that the (strongly) balanced forward degree sequence can be computed in polynomial time in the above cases.

The starting point for this topic was motivated by the following hypothetic situation:

> *A company has two open positions. On the waiting list there are n almost equally good candidates. Some pairs of the candidates can work together, some pairs can not. These pairs are known to the company. Most likely a person will accept the offer as soon as he gets it. It is also possible that he gets some better job and turns our offer down. We want to offer these candidates one at a time and get an immediate response. The goal is to maximize the likelihood of hiring two candidates that can work together. In what order the company should give the offers to the candidates?*

In some variants, the acceptance probabilities are close to 0 instead of close to 1, or the number of open positions is greater than two. (In this case, the company wants to hire candidates that can mutually work with each other.) We refer to all of these situations as *offer rejection problems.* Our goal was to find the best strategy dealing with the possible rejections.

The offer rejection problem lead to a subject that is interesting in its own right, namely, the *forward degree sequences of graphs.* A forward degree sequence arises from an ordering σ of the vertices of a graph. We eliminate the vertices according to this ordering, and the forward degree d_v^σ of a vertex v is its degree in the remaining graph when we eliminate it.

The idea of forward degree sequences is related to two classical topics in graph theory, namely, the *degree sequence* and *vertex elimination order*. The degree sequences of graphs are well characterized in [20], [21], and [22]. The vertex elimination order gives a nice characterization of chordal graphs. (See [25].) The forward degree sequences we define and study here arise from very different questions and are of different nature.

There are several nice connections between the offer rejection problem and the forward degree sequences. Here we focus on the pure graph-theoretical aspects. One connection arises when we associate to each forward degree sequence a polynomial $P_\sigma(z) = \sum_{v\in V} z^{d_v^\sigma}$. We find that the offer rejection problem with rejection probability q is equivalent to the problem of finding the σ which minimizes $P_\sigma(1/q)$ in the graph where edges represent pairs that can not work together. We will define an ordering σ to be more *balanced* than τ if $P_\sigma(1/q) \leq P_\tau(1/q)$ for every probability q. A related notion is *strongly balanced*. It is an interesting combinatorial question whether a graph has a most balanced (strongly balanced) forward degree sequence. We prove that this is true in some nice classes of graphs, for example, chordal graphs and 3-regular graphs, where we also give a polynomial time algorithm to find the most balanced sequence.

The forward degree sequences carry a lot of information of their graphs. One may easily express some usual graph parameters in terms of properties about the forward degree sequences. (See Section 4.2.) Here we define some new graph parameters (Section 4.3.3) based on the forward degree sequences. These parameters, besides their close relation to the offer rejection problem, are of pure graph-theoretic interest as well. One of the interesting problems that remains open is how to compute some of the parameters in polynomial time.

Our main result, that every 3-regular graph has a most balanced forward degree sequence, gives some new insight to the graph isomorphism problem ([23] and [24]).

4.1.1 Outline of the chapter

In Section 4.2 we define the forward degree sequences and their associated polynomials; we also give some simple properties of the sequences and polynomials, and their relation to some classical graph parameters. In Section 4.3 we define and study the balanced (strongly balanced) sequences. We define some graph parameters based on the forward degree sequences in Section 4.3.3. In Section 4.4 we consider the class of graphs that has a most balanced (strongly balanced) sequence. In Section 4.4.1 we use the closure property of these classes to deduce that some families of graphs, including all the forests and the chordal graphs, have most strongly balanced sequences. In Section 4.4.2, we prove our main theorem that every graph with degree at most 3 has a most balanced forward degree sequence.

4.1.2 Notations

In most situations, G or H represents graphs; S represents sequences; P and Q represent polynomials (in one variable); u, v, w, x, and y represent vertices in graphs; z will be used as the variable in polynomials; σ, ϕ, or π represents permutations (orderings).

All the graphs we consider will be undirected, unweighted, simple graphs. Let $G = (V, E)$ be a graph. We use the conventional notations in graph theory. Denote by $\Delta(G)$ the maximum degree in G, by $\delta(G)$ the minimum degree in G, by $\alpha(G)$ the size of a largest independent set in G, and by $\omega(G)$ the size of largest clique in G. The girth of G is the length of shortest cycle in G; it is defined to be ∞ if G is cycle free.

For any graph G, $E(G)$ is the set of edges in G, $V(G)$ is the set of vertices in G. The complement of G is denoted by $\overline{G}$. For any $S \subseteq V$, $G[S]$ is the subgraph of G induced by S. For any $v \in V$, the degree of v in G is denoted by d_v^G and the neighbors of v denoted by $N_G(v)$. The induced subgraph of G by deleting v is denoted by $G - v$. If $x, y \in V$, and xy is an edge in G, $G - xy$ is the graph G with edge xy deleted. If xy is not an edge in G, $G + xy$ is the graph gotten by adding the edge xy to G.

Let $S = (s_1, \cdots, s_n)$ be a sequence. We denote by $S(i)$ the i-th element of S. The concateneation of S and a new element v, $(s_1, \cdots, s_n, v)$, is denoted by (S, v); and the concateneation of a new element v and S, $(v, s_1, \cdots, s_n)$, is denoted by (v, S); $S[i \leftrightarrow j]$ is the sequence we get from S by exchanging the elements on the i-th position and the j-th position. We view a permutation on an n element set as a sequence of length n.

If S is a sequence of integers, we denote by $N_S(k)$ the number of occurrences of k in S, and denote by $\hat{S}$ the sorted list of S in non-increasing order.

We define the lexicographical order. If S_1 and S_2 are two sequences of integers, $S_1 \leq_{lex} S_2$ if $S_1 = S_2$ or there is an i such that $S_1(i) < S_2(i)$ and $S_1(j) = S_2(j)$ for all $j < i$.

4.2 Forward degree sequences

Definition Given a graph G and a permutation σ on the vertex set $\{v_1, v_2, \cdots, v_n\}$ of G, the *forward degree* of vertex $v = \sigma(i)$, denoted d_v^σ, is its degree in the induced subgraph $G' = G[\{\sigma(i), \sigma(i+1), \cdots, \sigma(n)\}]$. The *forward degree sequence* induced by σ, denoted S_σ, is the sequence $(d^\sigma_{\sigma(1)}, \ldots, d^\sigma_{\sigma(n)})$.

Notice that in our notation the graph G is implicitly indicated in σ. We eliminate the vertices of G according to the order σ, the forward degree of v is its degree in the remaining graph when we eliminate v. Another way to view this is that we direct all the edges of G according to σ. We direct edge uv from u to v if u comes before v in σ. Then the forward degree of a vertex is its out-degree in the directed graph. Related to our offer rejection problem, we associate to each permutation its forward degree sequence polynomial.

Definition For any sequence S of n non-negative integers, define P_S to be the polynomial $P_S(z) = \sum_{i=1}^n z^{S(i)} = \sum_{k=0}^\infty N_S(k) z^k$. Given a graph $G = (V, E)$ and a permutation σ of its vertices, the *forward degree sequence polynomial* induced by σ is $P_\sigma(z) := P_{S_\sigma}(z) = \sum_{v \in V} z^{d_v^\sigma} = \sum_{k=0}^\infty N_{S_\sigma}(k) z^k$.

Every two forward degree sequences π and σ of G have the same length (the number of vertices in G) and the same sum (the number of edges in G). The first fact implies that $P_\pi(1) = P_\sigma(1)$. The second fact implies that the derivatives satisfy $P'_\pi(1) = P'_\sigma(1)$. So we have

Proposition 4.2.1. *For any two forward degree sequences π and σ of G, the polynomial $P_\pi - P_\sigma$ is a multiple of $(z-1)^2$.*

The forward degree sequences of G carry a lot of graph theoretic information about G. As a warm up, we start by presenting a lemma where the forward degree sequences are related to the structure of the graph. Similar results will be shown later when we study 3-regular graphs.

Suppose a graph G has k connected components and σ is any permutation on vertices, then $N_{S_\sigma}(0) \geq k$ since the last vertex from each component in the order always has forward degree 0. Actually we can construct an ordering where the forward degree sequence contains exactly k zeros. The following lemma, which is a slightly stronger statement, is easy and we omit the proof.

Lemma 4.2.2. *If G is a connected graph and v is any vertex in G, then there is a permutation σ on the vertices of G such that $N_{S_\sigma}(0) = 1$ and the last vertex in the permutation is v.*

Definition Let $G = (V, E)$ be a graph. We define $\mathcal{S}_G$ to be the set of all the forward degree sequences, i.e., $\mathcal{S}_G = \{S_\sigma : \ \sigma \text{ is a permutation of vertices in } G\}$. Define $\hat{\mathcal{S}}_G$ to be the set $\{\hat{S} : S \in \mathcal{S}_G\}$. Define $\mathcal{P}_G$ to be the set of all the forward degree sequence polynomials.

The following list of obvious facts relate the forward degree sequences to several graph parameters. We omit the proofs.

Proposition 4.2.3. *For any graph G, the family $\mathcal{S}_G$ has the following properties. (a) $\Delta(G) = \max\{S(1) : S \in \mathcal{S}_G\}$. (b) $\delta(G) = \min\{S(1) : S \in \mathcal{S}_G\}$. (c) The number of connected components in G is the minimum occurrence of 0 in a single forward degree sequence, i.e., $\min\{N_S(0) : S \in \mathcal{S}_G\}$. (d) $\alpha(G)$ is the maximum number of 0's in a single forward degree sequence. (e) $\omega(G)$ is the maximum consecutively decreasing suffix of a single forward degree sequence, i.e., $\omega(G) = \max\{k : \exists S \in \mathcal{S}_G, S(n-i) =$*

i for all $0 \le i < k.\}$ *(f) The girth of* G *is the smallest length of a suffix in the form* $(2,1,1,\cdots,1,0)$ *of a single forward degree sequence, i.e.,* $\min\{k : \exists S \in \mathcal{S}_G, S(n-i) = 1$ *for all* $0 < i < k-1, S(n-k+1) = 2\}$, *or the girth is* ∞ *if such a suffix does not exist.*

4.3 Balanced and strongly balanced sequences

4.3.1 Balanced sequences

For any integers $n, m \ge 0$, let $\hat{\mathcal{S}}_{n,m}$ be the set of all non-increasing, non-negative integer sequences of length n where the sum of elements is m. We define a relation on $\hat{\mathcal{S}}_{n,m}$: $S_1 \precsim S_2$ if $P_{S_1}(z) \le P_{S_2}(z)$ for all $z \ge 1$. It is easy to check $\precsim$ is a partial order on the ordered sequences. If $S_1 \precsim S_2$, we also write $P_{S_1} \precsim P_{S_2}$. Thus we view $\precsim$ as a partial order on corresponding polynomials.

If G is a graph with n vertices and m edges, we have the induced partial orders $(\hat{\mathcal{S}}_G, \precsim)$ and $(\mathcal{P}_G, \precsim)$. Moreover, for any two orderings σ and π of vertices, we write $\sigma \precsim \pi$ and $S_\sigma \precsim S_\pi$ if $\hat{S}_\sigma \precsim \hat{S}_\pi$. $\precsim$ is no longer a partial order on all the permutations or on all the forward degree sequences of G, but it is still transitive and reflexive. We have an equivalence relation $\sigma \sim \pi$ if $\sigma \precsim \pi$ and $\pi \precsim \sigma$. (Similarly for the forward degree sequences.) Clearly σ and π are equivalent if and only if $\hat{S_\sigma} = \hat{S_\pi}$.

Definition A graph G is called *balanced* if there exist a minimum element in $(\hat{\mathcal{S}}_G, \precsim)$. The class of all balanced graphs is denoted by $\mathcal{B}$.

As a partial order on a finite set, $(\hat{\mathcal{S}}_G, \precsim)$ and $(\mathcal{P}_G, \precsim)$, always have minimal elements. For any permutation σ, we call σ and S_σ $\precsim$ minimal if $\hat{S}_\sigma$ is minimal. If there exists a permutation σ of vertices such that $\sigma \precsim \pi$ for any permutation π of vertices, we call σ a $\precsim$ *minimum ordering* of G, and S_σ a $\precsim$ *minimum forward degree sequence.*

4.3.2 Strongly balanced sequences

We introduce another relation which is a sufficient condition for checking $\precsim$. Let S be a sequence in $\hat{\mathcal{S}}_{n,m}$. If a and b both appear in S and $a \ge b+2$, the concentration operation $(a, b) \to (a-1, b+1)$ is performed by changing one a and one b to an $a-1$ and a $b+1$, then sort the sequence in non-increasing order. We define $S_1 \precsim_S S_2$ if we can get S_1 from S_2 by 0 or more steps of concentration. In this case we also write $P_{S_1} \precsim_S P_{S_2}$. If σ and π are two permutations on the vertices of a graph G, we write $\sigma \precsim_S \pi$ and $S_\sigma \precsim_S S_\pi$ if $\hat{S}_\sigma \precsim_S \hat{S}_\pi$ (and $P_\sigma \precsim_S P_\pi$). It is easy to see $S_1 \precsim S_2$ if $S_1 \precsim_S S_2$; $\precsim$ are partial orders on $\hat{\mathcal{S}}_{n,m}$, $\hat{\mathcal{S}}_G$, and $\mathcal{P}_G$.

Definition A graph G is called *strongly balanced* if there exist a minimum element in $(\hat{\mathcal{S}}_G, \precsim_S)$. The class of all the strongly balanced graphs is denoted by $\mathcal{B}_S$.

If G is strongly balanced, any permutation σ when $\hat{S}_\sigma$ is minimum is called a $\precsim_S$ *minimum ordering*, S_σ is called a $\precsim_S$ *minimum forward degree sequence*, or a *most strongly balanced forward degree sequence*. If $\hat{S}_\sigma$ is $\precsim_S$ minimal, we also call σ or S_σ $\precsim_S$ minimal.

As an example, we note that every tree has a most strongly balanced forward degree sequence $(1, 1, \cdots, 1, 0)$, which is $\precsim_S$ minimum even in the whole $\hat{\mathcal{S}}_{n,n-1}$. Similarly, every forest is in $\mathcal{B}_S$. (See Corollary 4.4.3 for a generalization of this fact.)

Next, we give several characterizations of the relation $\precsim_S$.

Proposition 4.3.1. *Let S_1 and S_2 be two sequences in $\hat{\mathcal{S}}_{n,m}$. $S_1 \precsim_S S_2$ if and only if $P_{S_2} - P_{S_1} = (z-1)^2 Q$ where Q is a polynomial in z with positive integer coefficients.*

Sketch of Proof. The *only if* part is trivial. The *if* part may be proved by induction on the sum of coefficients in Q. The key observation is that for any "segment" (a, b) in Q ($b \le a$ and the coefficient of z^i in Q is positive for each $b \le i \le a$, while the coefficient of z^{a+1} and z^{b-1} are both 0), we could change S_2 to S' by $(a+2, b) \to (a+1, b+1)$ and $P_{S'} - P_{S_1} = (z-1)^2 Q'$, where $Q' = Q(z) - (z^b + \ldots + z^a)$ is positive. □

Suppose S_1 and S_2 are two ordered sequences of length n. Define $a_i := S_2(i) - S_1(i)$. Define $s_0 = 0$ and $s_i = s_{i-1} + a_i$ for $1 \le i \le n$. Clearly s_i is the difference between the sum of i-prefix in S_1 and S_2. We may draw the points (i, s_i) in the plane, connect each (i, s_i) to $(i+1, s_{i+1})$ by a straight line to make a ploynomial path. We call this the *walk of* $S_1 - S_2$. Clearly the walk always starts from the x-axis and returns to x-axis at the end. Notice that, in any concentration step, the sum of the first i elements will not increase. On the other hand, if the walk of $S_1 - S_2$ does not go below the x-axis, we can concentrate from S_1 to S_2, each step takes $(S_1(i), S_1(j)) \to (S_1(i)-1, S_1(j)+1)$ where i and j are the end points of a segament strictly above x-axis in the walk. Thus we get

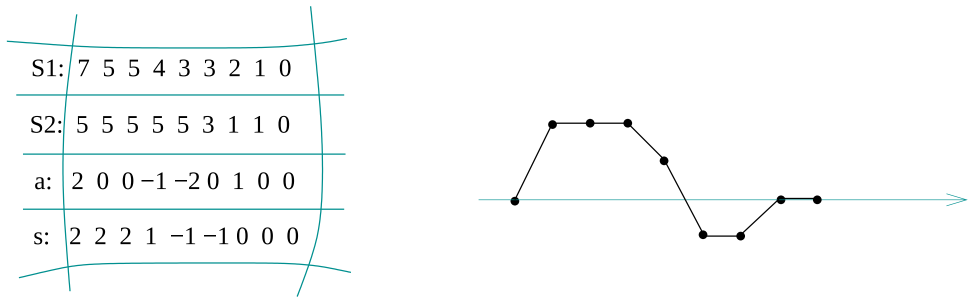

Figure 4.1: Two sequences and the polygonal walk.

Proposition 4.3.2. *Let $S_1, S_2 \in \hat{\mathcal{S}}_{n,m}$ for some n and m. The following are equivalent. (a). $S_2 \precsim_S S_1$. (b). For any i, the sum of the first i elements in S_1 is no less than the sum of the first i elements in S_2. (c). The walk of $S_1 - S_2$ does not go below x-axis.*

4.3.3 Finding the minimal sequences

Given a graph G, let S^* be the lexicographically smallest sequence in $\hat{\mathcal{S}}_G$. When z is big enough, $P_{S^*}(z) < P_S(z)$ for any other $S \in \hat{\mathcal{S}}_G$. Therefore, S^* is one of the minimal elements in both the orders $\precsim$ and $\precsim_S$. If $G \in \mathcal{B}$ or $G \in \mathcal{B}_S$, we must have S^* as the minimum sequence. From S^* we define some interesting graph parameters.

Definition Let G be a graph and S^* be the smallest lexicographical sequence in $\hat{\mathcal{S}}_G$. Define $m(G) := S^*(1)$, i.e., the largest forward degree in S^*. Define $N_k(G) := N_{S^*}(k)$ to be the number of occurrences of k in S^*. And define $N(G) := N_{m(G)}(G) = N_{S^*}(m(G))$.

Clearly $m(G)$, $N(G)$, and $N_k(G)$ are invariant under graph isomorphism, for all k.

Given a graph G, $m(G)$ is computable in polynomial time by the following algorithm: We start from the empty sequence; find a vertex v in G with the smallest degree; put v

as the next element in our sequence and delete v from G. Iterate this until G is empty. Thus we get an ordering of the vertices and a forward degree sequence. We claim $m(G)$ is the largest number in the sequence. Indeed, the correctness of the algorithm follows easily from the following observations:

Fact 4.3.3. *Let G' be a subgraph of G, then $m(G') \leq m(G)$.*

Fact 4.3.4. *Let G be a graph, v a vertex in G with the minimum degree, and $G' = G - v$. Then $m(G) = \max\{d_v, m(G')\}$.*

We do not know if $N(G)$ is computable in polynomial time. It seems to be a very hard problem.

If $G \in \mathcal{B}_S$, the most strongly balanced forward degree sequence must be S^*. In finding S^*, we are trying to minimize the number of occurrences of $m(G)$ in a forward degree sequence. On the other hand, by Proposition 4.2.3(c) and Proposition 4.3.2, the most strongly balanced sequence must contain as few 0's as possible, so the number of 0's in the sequence is exactly the number of connected components in G. With these observations, we can show there are graphs not in $\mathcal{B}_S$ and not even in $\mathcal{B}$. The graph in Figure 4.2 is not in $\mathcal{B}_S$. For the graph in Figure 4.3, S^* is the sequence $(3,3,3,3,3,3,3,3,2,2,2,2,1,1,1,0,0,0)$. It is not the $\precsim$ minimum sequence. Indeed, let σ be $(u_1,\cdots,u_5,v_1,v_2,v_3,u_6,\cdots,u_{15})$. $\hat{S}_\sigma$ is the sequence $(4,3,3,3,3,3,2,2,2,2,2,2,1,1,1,1,0,0)$. $P_\sigma(z) < P_{S^*}(z)$ when $1 < z < (\sqrt{5}+1)/2$. So the graph is not in $\mathcal{B}$.

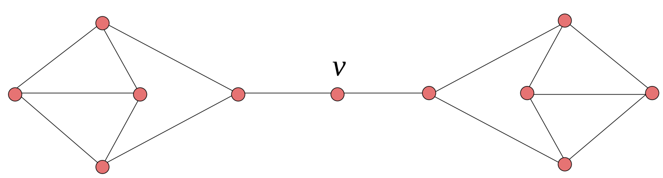

Figure 4.2: A graph that is not in $\mathcal{B}_S$.

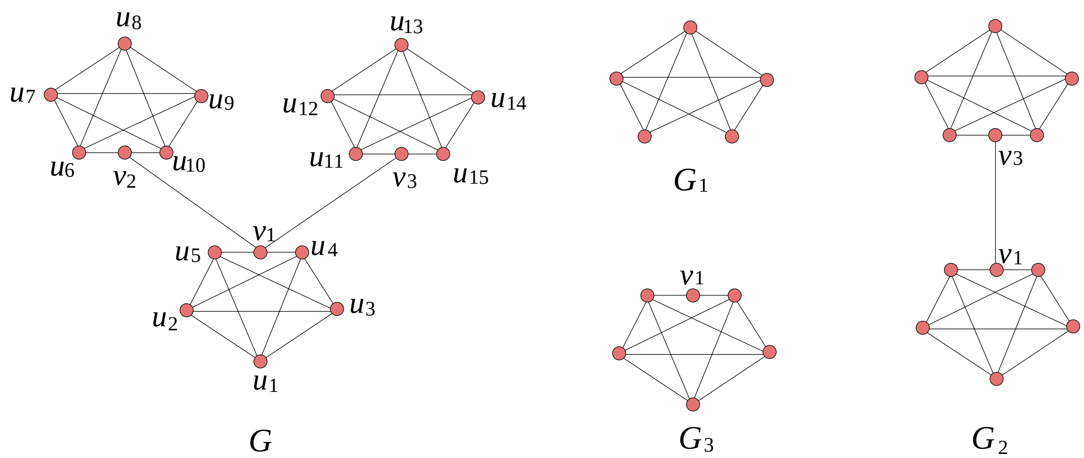

Figure 4.3: A graph G that is not in $\mathcal{B}$. Its subgraphs $G_1 \in \mathcal{B}_S$, $G_2 \in \mathcal{B}$, and $G_3 \in \mathcal{B}_S$.

4.4 Graphs inside $\mathcal{B}$ or $\mathcal{B}_S$

4.4.1 Closure properties

Given a graph $G \in \mathcal{B}_S$, we may construct a new graph G' from G by adding a new vertex v and connecting it to some vertices in G. We prove that the new graph is still in $\mathcal{B}_S$ if we do any of the following on G and v: (1) connect v to every vertex in G; (2) connect v to a clique in G; (3) connect v to two vertices x and y in G, where x and y are connected in G. (1) and (2) are also true if we replace $\mathcal{B}_S$ by $\mathcal{B}$. We prove these in the following propositions.

Proposition 4.4.1. *If $G = (V, E) \in \mathcal{B}$ and $G' = (V \cup \{v\}, E \cup \{vw | w \in V\})$, then $G' \in \mathcal{B}$. Moreover, if $G \in \mathcal{B}_S$, then $G' \in \mathcal{B}_S$. If σ is a $\precsim$ ($\precsim_S$) minimum ordering of the vertices of G, then $\sigma' = (\sigma, v)$ is $\precsim$ ($\precsim_S$) minimum for G'.*

Proposition 4.4.2. *If $G = (V, E) \in \mathcal{B}$, K is a clique in G, and $G' = (V \cup \{v\}, E \cup \{vw | w \in K\})$, then $G' \in \mathcal{B}$. Moreover, if $G \in \mathcal{B}_S$, then $G' \in \mathcal{B}_S$. If σ is a $\precsim$ ($\precsim_S$) minimum ordering of the vertices of G, then $\sigma' = (v, \sigma)$ is $\precsim$ ($\precsim_S$) minimum for G'.*

The proofs of the two propositions above are consequences of an easy shifting argument.

Proposition 4.4.2 immediately gives the fact that there is a most strongly balanced sequence for any chordal graph. First we recall some definitions and facts from graph theory([25], p. 198 to p. 200). A *chord* of a cycle C is an edge not in C that has endpoints in C. A graph G is *chordal* if every cycle of length at least 4 has a chord. A vertex of G is *simplicial* if its neighborhood in G induces a clique. A *simplicial elimination ordering* is an order $v_n, \ldots, v_1$ in which vertices can be deleted so that each vertex v_i is a simplicial vertex of the remaining graph induced by $\{v_1, \ldots, v_i\}$. It is well known that a graph G is chordal if and only if it has a simplicial elimination ordering.

Corollary 4.4.3. *(of Proposition 4.4.2) The family of chordal graphs is contained in* $\mathcal{B}_S$.

Corollary 4.4.4. *(of Proposition 4.4.2) Let G be a chordal graph. All simplicial elimination orderings of G give the same multiset of the forward degree sequences.*

Proposition 4.4.5. *If $G = (V, E) \in \mathcal{B}_S$, x and y are two connected vertices in G, and $G' = (V \cup \{v\}, E \cup \{vx, vy\})$, then $G' \in \mathcal{B}_S$. If σ is a $\precsim_S$ minimum ordering of the vertices of G, then $\sigma' = (v, \sigma)$ is $\precsim_S$ minimum for G'.*

Proof. Without loss of generality, we may assume G is connected. (If G is not connected, we may simply consider every connected component of G.) Therefore, the number 0 appears exactly once in the forward degree sequence induced by σ. Clearly $P_{\sigma'}(z) = P(z) + z^2$.

Consider any ordering π' of G'. Without loss of generality, x comes before y in π'. By deleting v in π' we get π, an ordering of G. $P_\pi(z) - P(z) = (z-1)^2 Q(z)$, where Q is a polynomial in z with positive integer coefficients. Let $a = d_x^\pi$ and $b = d_y^\pi$.

Based on the position of v, x, and y in π', we have 3 cases.

(a) If v comes before x and y, then $d_v^\sigma = 2$ and $d_w^\sigma = d_w^{\sigma'}$ for any w in G. So $P_{\pi'}(z) = P_\pi(z) + z^2$. $P_{\pi'}(z) - P_{\sigma'}(z) = P_\pi(z) - P(z) = (z-1)^2 Q(z)$.

(b) If v comes between x and y, $P_{\pi'}(z) = P_\pi(z) + z + z^{a+1} - z^a$. $P_{\pi'}(z) - P_{\sigma'}(z) = P_\pi(z) - P(z) + z - z^2 + z^{a+1} - z^a = (z-1)^2 Q(z) + (z-1)(z^a - z)$. If $a \neq 0$, $(P_{\pi'}(z) - P_{\sigma'}(z))/(z-1)^2$ is a polynomial with positive integer coefficients. Otherwise, x has forward degree 0 in

π, yet x is not the last vertex. So π contains at least 2 vertices of forward degree 0. Therefore, the constant term of Q is positive, and $P_{\pi'}(z) - P_{\sigma'}(z) = (z-1)^2(Q(z)-1)$, where all the coefficients of $Q(z)-1$ are positive integers.

(c) If v comes after x and y, $P_{\pi'}(z) = P_{\pi}(z)+1+z^{a+1}-z^a+z^{b+1}-z^b$. $P_{\pi'}(z)-P_{\sigma'}(z) = (z-1)^2Q(z)+(z-1)(z^a-z+z^b-1)$. The analysis is exactly the same as in the previous case.

In each case, $(P_{\pi'} - P_{\sigma'})/(z-1)^2$ has positive coefficients. Using proposition 4.3.1, σ' is $\precsim_S$ minimum for G'. □

4.4.2 Graphs with low degrees

We denote by $\mathcal{D}_k$ the class of graphs with all degrees at most k: $\mathcal{D}_k = \{G : \Delta(G) \leq k\}$. Any graph in $\mathcal{D}_2$ is a disjoint union of paths and cycles. By the propositions in Section 4.4.1, $\mathcal{D}_2 \subset \mathcal{B}_S$. By the examples in Figure 4.2 and Figure 4.3, we have $\mathcal{D}_3 \not\subseteq \mathcal{B}_S$ and $\mathcal{D}_4 \not\subseteq \mathcal{B}$. In the rest of this section we prove that $\mathcal{D}_3 \subset \mathcal{B}$, especially the class of 3-regular graphs is contained in $\mathcal{B}$.

Biconnectivity and block structures

For the graphs in $\mathcal{D}_3$, it turns out that the biconnectivity and the blocks of a graph play a crucial role in finding the most balanced sequence. We recall some definitions and facts about biconnected graphs and blocks.

Definition A graph is called *biconnected* if it is connected, has at least 3 vertices and contains no cut point. A maximal connected subgraph that has no cut point is called *a block*

In the standard approach the blocks define a partition of the edges of a graph. Blocks form a cactus-like structure and the so called block-cutpoint graph is a tree. In our approach we use a slightly different decomposition.

Definition Given a graph G. A maximal biconnected subgraph of G is called *a cluster*. A vertex which does not belong to any biconnected subgraph is called a *router*. A *room*

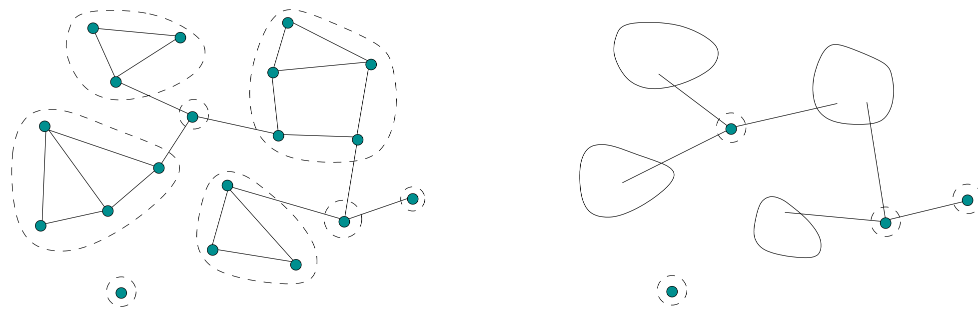

Figure 4.4: Rooms, clusters, routers, and a building map.

is either a cluster or a router.

Clearly, a router is a vertex that does not belong to any cycle. Either it is an isolated point, or every block containing it is a bridge. The following facts are easy consequences of standard properties of the block decomposition of a graph. (See [19] and [25].)

Fact 4.4.6. *Let G be a graph in $\mathcal{D}_3$; let R_1 and R_2 be any two distinct rooms in G. Then,*

(a) R_1 and R_2 do not share any common vertex.

(b) There is at most one edge in G between the vertex set of R_1 and the vertex set of R_2.

Thus, for graphs in $\mathcal{D}_3$, the rooms form a partition of vertices.

Definition Let $G \in \mathcal{D}_3$ with rooms $R_1, \cdots, R_k$. The *building map* G^R is the graph with vertex set $\{R_1, \cdots, R_k\}$ and R_iR_j is an edge in G^S if and only if there is an edge between R_i and R_j in G.

Fact 4.4.7. *For any $G \in \mathcal{D}_3$, the graph G^S is a forest. The number of connected components in G^S is the number of connected components in G.*

Biconnected graph in $\mathcal{D}_3$

If G is a connected graph in $\mathcal{D}_d$, and there is at least one vertex with degree smaller than d, then we can always find a forward degree sequence without any forward degree d. Indeed, we can pick any vertex of degree less than d, and notice that in the new graph there is at least one vertex of degree less than d in each of its connected component. Thus we have,

Proposition 4.4.8. *For any connected graph G with maximum degree d and minimum degree less than d, we have $m(G) < d$.*

Corollary 4.4.9. *For any d-regular graph G, $m(G) = d$ and $N(G)$ equals the number of connected components in G.*

We investigate the biconnected graphs in $\mathcal{D}_3$. For these graphs, we have a similar yet stronger statement.

In the proof of the next Lemma and in the rest of this section, we call a forward degree sequence *x-y-good* if it starts with x, ends with y, avoids any 3, and contains only one 0.

Lemma 4.4.10. *Suppose $G = (V, E)$ is a biconnected graph in $\mathcal{D}_3$. For any $v_s \in V$ such that $d^G_{v_s} = 2$ and any $v_f \in V - v_s$, there exists an ordering σ of vertices such that it starts with v_s, ends with v_f, and the forward degree sequence obtained by σ on G satisfies $N_{S_\sigma}(3) = 0$ and $N_{S_\sigma}(0) = 1$, i.e., it avoids 3 and contains only one 0.*

Proof. The proof is by induction on n, the number of vertices in G. The base case, when $n = 3$, is trivial. For n greater than 3, $G' = G - v_s$ is a connected graph in $\mathcal{D}_3$ with $n - 1$ vertices. Suppose x and y are the two neighbors of v_s, in G' we have $1 \le d^{G'}_x, d^{G'}_y \le 2$. We discuss two possible cases.

1. G' is biconnected.

In this case, $d^{G'}_x = d^{G'}_y = 2$ and at least one of them, say x, is not v_f. $\sigma = (v_s, \sigma')$ is v_s-v_f-good in G , where σ' is x-v_f-good in G'.

2. G' is not biconnected.

In this case, there should exist v_c which is a cut point of G'. Since the degree of v_s in G is 2 and G is biconnected, $G' - v_c$ has exactly two connected components, say, $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$, and v_s has neighbor in both components. We assume $x \in V_1$ and $y \in V_2$.

Based on H_1 and H_2, we define two new graphs. $G_1 = G[V_1 \cup \{v_c, v_s\}] + v_s v_c$, $G_2 = G[V_2 \cup \{v_c, v_s\}] + v_s v_c$. It is easy to show that they are biconnected graphs in $\mathcal{D}_3$ with at least three and less than n vertices, and v_s is a degree 2 vertex in both G_1 and G_2.

Now we construct a v_s-v_f-good ordering in G. Without loss of generality, v_f is in G_1. By induction, there is a v_s-v_f-good ordering (v_s, σ_1, v_f) in G_1, and a v_s-v_c-good ordering (v_s, σ_2, v_c) in G_2. Let $\sigma = (v_s, \sigma_2, \sigma_1, v_f)$. It is routine to check that σ is v_s-v_f-good in G. □

Now we prove that a v_s-v_f-good forward degree sequence, for any v_s and v_f with the degree of v_s being 2, is a most strongly balanced sequence for the graph.

Corollary 4.4.11. *Any biconnected graph with maximum degree* 3 *and minimum degree less than* 3 *belongs to the class* $\mathcal{B}_S$.

Proof. Let σ be a v_s-v_f-good ordering for any v_s and v_f with the degree of v_s being 2. Consider any other ordering π and the maximum number d appearing in S_π.

Since the graph is biconnected, $d > 1$. If $d = 2$, then $P_\pi(z) - P_\sigma(z) = (z-1)^2 c$ for some integer c. Since S_σ contains only one 0, the constant term of $P_\pi(z) - P_\sigma(z)$ is non-negative. If $d = 3$, then $P_\pi(z) - P_\sigma(z) = (z-1)^2(az + b)$ for some integer $a > 0$ and b. Again, since S_σ contains only one 0, the constant term of $P_\pi(z) - P_\sigma(z)$ is non-negative. In either case, $\sigma \precsim_S \pi$ by Proposition 4.3.1. □

In the proof of Lemma 4.4.10, we actually outlined an algorithm to find a good sequence. The complexity of the algorithm is easily analyzed. We have

Proposition 4.4.12. *Given any biconnected graph with n vertices, and with maximum degree* 3 *and minimum degree less than* 3*, the most strongly balanced (ordered) forward degree sequence is computable in* $O(n^3)$ *time.*

$\mathcal{B}$ contains $\mathcal{D}_3$

Lemma 4.4.13. *Any connected graph G with maximum degree* 3 *and minimum degree less than* 3 *belongs to the class $\mathcal{B}$. Moreover, the $\precsim$ minimum forward degree sequence is computable in polynomial time.*

Proof. We call a vertex *loose* if its degree is less than 3. We analyze the clusters in G and the building map G^S. We call a cluster R *bad* if it is a leaf in G^S and all its vertices have degree 3 in G. Let b be the number of bad clusters.

If we want to avoid the forward degree 3, each of the bad clusters contributes at least one 0. We have

Fact 4.4.14. *In any forward degree sequence of* G *that avoids* 3*, the number of* 0*'s is at least* $\max\{1, b\}$*.*

On the other hand, we can always achieve the minimum possible number of 0's. Here we sketch the procedure: Pick any loose vertex x and find its room R_x. We view G^S as a rooted tree with root R_x. For any leaf L which is not a bad cluster, by Lemma 4.4.10, we can eliminate it without producing any 3 or 0, or changing the number of bad clusters. We repeat this until all the leaves are bad. Now, we find an order from the root down to the leaves. We start from x. For any non-leaf room R, we start from a loose vertex, and eliminate its vertices according to the ordering provided by Lemma 4.4.10 which ends in any of its ports to R's children of R. Thus we do not have any forward degree 3 or 0, and created a loose vertex for each of R's children. Finally we have d leaves in G^S, accordingly d connected components in G each has a loose vertex. By Lemma 4.4.10, we finish by d orderings containing one 0 each.

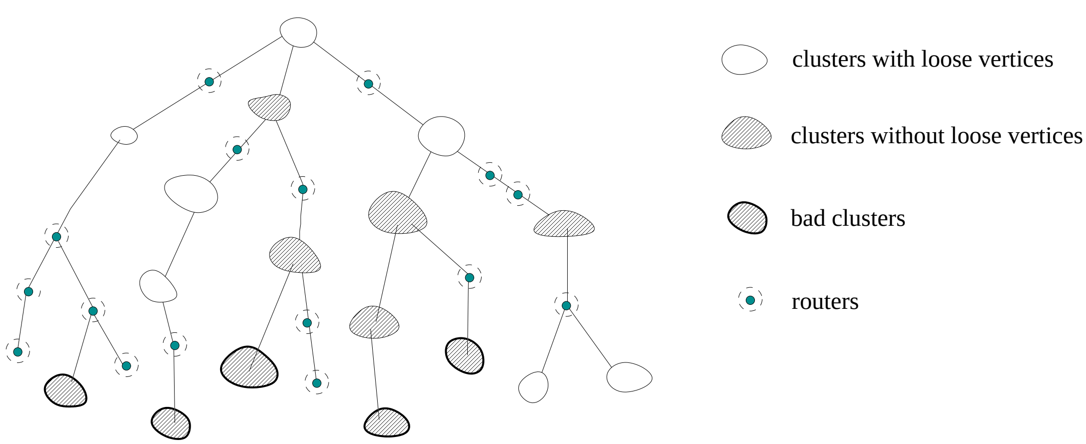

Figure 4.5: The good leaves are treated bottom up, then the non-leaf nodes are treated top down.

Fact 4.4.15. *There is a forward degree sequence of G that avoids* 3 *and contains* $\max\{1, b\}$ 0*'s.*

Now we finish the proof of Lemma 4.4.13 by showing that any sequence σ provided by the above claim is $\precsim$ minimum for G. We prove this by induction on the size of G. The base case is trivial. Assume for any graph with less vertices there is a $\precsim$ minimum ordering that has as few 0's as possible under the condition that 3 does not appear.

Consider any other ordering π. We consider two cases.

1. π avoids the forward degree 3. Then, $P_\pi(z) - P_\sigma(z) = (z-1)^2 c$ for some integer c. Because σ has the fewest number of 0's among all forward degree sequence without 3, $c \geq 0$ and hence $\sigma \precsim \pi$.

2. 3 appears in π as the forward degree for some v. We may assume it appears as the beginning and, by the inductive hypothesis, $\pi = (v, \pi')$, where π' is the $\precsim$ minimum sequence of $G - v$. So, 3 appears exactly once in S_π. $P_\pi(z) - P_\sigma(z) = (z-1)^2(z-c)$ for some integer c.

Our goal is to show that $c \leq 1$, or equivalently, that $N_{S_{\pi'}}(0) \geq N_{S_\sigma}(0) - 1$. If there is no bad clusters in G, then $N_{S_\sigma}(0) = 1$ and $N_{S_{\pi'}}(0) \geq N_{S_\sigma}(0) - 1$. Otherwise, there are $b \geq 1$ bad clusters. $G - v$ contains at least $b - 1$ bad clusters, so by fact 4.4.14, $N_{S_{\pi'}}(0) \geq b - 1 = N_{S_\sigma}(0) - 1$.

The procedure we outlined in this proof gives a best forward degree sequence in polynomial time, provided the algorithm in Proposition 4.4.12 as a sub-routine. □

Now we are ready to prove the main result of this section.

Theorem 4.4.16. *Any graph with all degrees at most* 3 *has a* $\precsim$ *minimum forward degree sequence. There is a polynomial time algorithm to compute the* $\precsim$ *minimum forward degree sequence. Finally, the class of* 3*-regular graphs is contained in* $\mathcal{B}$*.*

Proof. We may assume the graph G is connected. If there is a vertex with degree less than 3, the statement is true by Lemma 4.4.13. Otherwise, G is 3-regular, with vertices $v_1, v_2, \cdots, v_n$. Let $G_i = G - v_i$, each connected component of G_i has a vertex of degree less than 3. Therefore, by Lemma 4.4.13, there is a $\precsim$ minimum forward degree sequence σ'_i for G_i. Let $\sigma_i = (v_i, \sigma'_i)$, then $P_{\sigma_i}(z) = z^3 + P_{\sigma'_i}(z)$.

For any ordering (v_i, σ'), $(v_i, \sigma'_i) \precsim (v_i, \sigma')$. So we only need to show that there is a $\precsim$ minimum polynomial among $P_{\sigma_1}, \cdots, P_{\sigma_n}$.

Actually $\precsim$ is a linear order on the set of $\{P_{\sigma_i} : 1 \leq i \leq n\}$. For any i and j, $S_{\sigma'_i}$ and $S_{\sigma'_j}$ do not contain any forward degree larger than 2, since they are $\precsim$ minimum sequences. So,

$$P_{\sigma'_i}(z) - P_{\sigma'_j}(z) = P_{\sigma_i}(z) - P_{\sigma_j}(z) = (z-1)^2 c$$

for some integer c. That is, they are $\precsim$ ($\precsim_S$) comparable. □

Chapter 5

Santa Claus' Towers of Hanoi

5.1 The towers of hanoi and two new variations

The well known Towers of Hanoi, treasured by most mathematicians and many others as their childhood toy for mental gymnastics, was invented by the French mathematician Edouard Lucas in 1883. In the legend, 64 sacred disks were initially stacked in increasing size on one of three pegs, with the largest at the bottom. A monk has to move the entire tower to another peg. The disks are fragile; only one can be carried at a time. And most importantly, the monk must obey the following rule:

The Divine Rule: A disk may not be placed on top of a smaller disk.

The disks are of size $1, 2, \ldots, n$, usually we denote the disk of size i by $\langle i \rangle$; and we denote the set of disks by $[n] = \{\langle i \rangle : 1 \leq i \leq n\}$. The three pegs are P_1, P_2, and P_3.

The Towers of Hanoi and its many variations are well studied. Here we just mention a few of them: The cyclic towers of Hanoi was first studied by Atkinson [26]. In [32], Klein and Minsker solved the variation where there can be bigger disks above smaller disks in the initial configuration, but the moves still obey the divine rule. The multi-peg Towers of Hanoi problem was proposed by Stewart [34], and remains a big open problem. (See related works by Stewart [35], Frame [30], Szegedy [37], Klavžar, Milutinović, and Petr [31], and Chen and Shen [27].) For a good bibliography with more than 200 entries on this subject, see Paul Stockmeyer's manuscript ([36]), which can be found on the Internet.

Here we study two new variations of the game. Both are on three pegs, and the monk may violate the divine rule slightly in some way.

In the *sinner's mode*, the divine rule may be violated a certain number of times. There is a prescribed number k. And the rule becomes

> *Sinner's Rule with k cheats*: For at most k moves in the whole process, some disk may be placed directly on the top of a smaller disk.

In the *Santa Claus'* mode we have, maybe for a better aesthetic reason, another slightly imperfect rule. There is a number d. And the rule is

> *Santa's Rule with discrepancy d*: When a disk $\langle x \rangle$ is put on a pile of other disks, before the move, disks smaller than $\langle x \rangle$ can only occur at the top up to $d - 1$ positions in that pile, but none of them can be of size less than or equal to $x - d$.

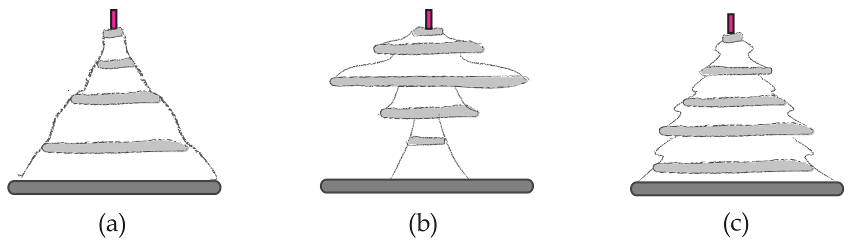

Figure 5.1: (a) a peg in the classical Hanoi Towers (b) a peg in the Sinner's mode (c) a peg in Santa Claus' mode.

We justify the two new versions based on two reasons. In our age, not so many people care about pure divine rules. Besides, Santa Claus likes Christmas trees. It is interesting to point out that when E. Lucas wrote about the towers of Hanoi, he used the name N. Claus.

After we searched through the literature, we believe the sinner's tower has not been studied before. In Section 5.3 we give a procedure for this problem and prove the optimality of our procedure. In the literature we found a variant that is very similar to Santa's tower in [38]. We reword it here as the *generous Santa Claus' mode*.

> *Generous Santa's Rule with discrepancy d*: When a disk $\langle x \rangle$ is put on a pile of other disks, smaller disks may occur at any positions in that pile, but none of them can be of size less than or equal to $x - d$.

The solution to the Santa Claus' mode and the generous Santa Claus' mode are almost identical. We give the complete solution to them in Section 5.4, thus resolve the open question raised by D. Wood in [38]. We use the last section of this chapter to present the story of our work. Especially, D. Poole published a solution to Wood's problem in 1992 [33]. We note the error in Poole's proof.

In the next section, we give notations, and prove some basic facts about Hanoi towers in the canonical version and our versions.

5.2 The hanoi towers

The classical game of moving a set X of disks from peg P to peg Q, where the disks are sorted in increasing order in both initial and final configurations, is denoted by $H(X, P, Q)$. The games under the sinner's rule, Santa's rule, and generous Santa's rule are denoted by $H_S(X, P, Q, k)$, $H_C(X, P, Q, d)$, and $H_G(X, P, Q, d)$, respectively. Often, we denote a move of disk $\langle x \rangle$ from P to Q as $\langle x \rangle(P \to Q)$.

A *state* is a specification of the sets of disks on three pegs, as well as their order on each peg. A *procedure* is a sequence of disk moves. Note that in our games there might be disks on the top of smaller disks in a state, and the moves in a procedure do not necessarily obey the divine rule.

Below we provide the canonical recursive procedure `hanoi` to solve the game $H([n], P_i, P_j)$, where $\{i, j\} \subset \{1, 2, 3\}$, $i \neq j$ and list some easy facts.

Procedure 5.2.1. `hanoi([n], a, b)`

```
c := 6-a-b;
hanoi([n-1], a, c);
move disk n from peg a to peg b;
hanoi([n-1], c, b);
```

Fact 5.2.2. *The procedure* `hanoi([n], a, b)` *obeys the divine rule, takes* $2^n - 1$ *moves, and the i-th biggest disk, i.e.,* $\langle n - i + 1 \rangle$*, is moved* 2^{i-1} *times.*

Fact 5.2.3. *Any procedure obeying the divine rule needs to take at least* $2^n - 1$ *moves, if the* n *disks are in the same peg at the end and the biggest disk is moved at least once.*

Definition Let $\mathcal{A}$ be a procedure with a sequence of m moves,

$$(\langle x_i\rangle(P_i \to Q_i) : i = 1 \cdots m).$$

The *reverse* of $\mathcal{A}$ is defined to be the sequence

$$(\langle x_{m+1-i}\rangle(Q_{m+1-i} \to P_{m+1-i}) : i = 1 \cdots m).$$

Lemma 5.2.4. *Let $\mathcal{A}$ be any procedure that takes a state S_1 to state S_2; let $\mathcal{A}'$ be the reverse of $\mathcal{A}$. Then $\mathcal{A}'$ takes S_2 to S_1. Moreover, if the moves of $\mathcal{A}$ obey the divine rule, yet there are k disks in S_1 such that each one is bigger than the one below it, then $\mathcal{A}'$ violates the divine rule at most k times.*

Proof. It is obvious that $\mathcal{A}'$ takes the state S_2 back to S_1. Suppose there are m moves in $\mathcal{A}$. If $\langle x_{m+1-i}\rangle(Q_{m+1-i} \to P_{m+1-i})$ is a move in $\mathcal{A}'$ which violates the divine rule, then the move $\langle x_i\rangle(P_i \to Q_i)$ in $\mathcal{A}$ takes $\langle x\rangle$ away from a smaller disk below it. Since there are at most k such disks in the initial state of $\mathcal{A}$, and $\mathcal{A}$ never violates the divine rule, the lemma follows. □

Definition Let $\mathcal{A}$ be a procedure playing any game of towers of Hanoi with disk set X; let D be any subset of X; the cost of D in $\mathcal{A}$, $c(\mathcal{A}, D)$, is defined to be the number of steps $\mathcal{A}$ moves disks in D.

5.3 The sinner's towers

Define $S(n, k)$ to be the minimum number of moves one needs to move n disks obeying the sinner's rule. It is clear that $S(n, k) = 2n - 1$ when $k \geq n - 2$.

Definition We define a function $g(n, k)$ on integers to be

$$g(n,k) = \begin{cases} 2n - 1 & \text{if } n \leq k + 2, \\ 4n - 2k - 5 & \text{if } k + 2 \leq n \leq 2k + 2, \\ 2^{n-2k} + 6k - 1 & \text{if } 2k + 2 \leq n. \end{cases}$$

Note that in the definition the values are consistent where the ranges overlap.

Proposition 5.3.1. *For all n and k, $S(n,k) \le g(n,k)$.*

Proof. For $n > 2k+2$, we do the game $G = H_S([n], P_1, P_2, k)$ in five stages.

(a). We simulate the game $H([n-2k-1], P_1, P_3)$ to move the top $n-2k-1$ disks to the third peg.

(b). Now we have $2k+1$ disks left on P_1. We divide the top $2k$ disks into k pairs of adjacent disks. Starting from the top, for each pair $\{\langle i\rangle, \langle i+1\rangle\}$, we move $\langle i\rangle$ to P_2, then move $\langle i+1\rangle$ to P_3 (thus violate the divine rule once), then move $\langle i\rangle$ to P_3.

(c). Move $\langle n\rangle$ to P_2.

(d). For each pair $\{\langle i\rangle, \langle i+1\rangle\}$ in (2), we move $\langle i\rangle$ to P_1, then move $\langle i+1\rangle$ to P_2, and move $\langle i\rangle$ to P_2.

(e). Simulate the game $H([n-2k-1], P_3, P_2)$.

The total number of steps is

$$(2^{n-2k-1}-1) + 3k + 1 + 3k + (2^{n-2k-1}-1) = 2^{n-2k} + 6k - 1.$$

And the divine rule is violated k times in stage (2).

For $2k+2 \ge n > k+2$, we do the game G in five stages. Let $s = n-k-2$.

(a). Move $\langle 1\rangle$, $\langle 2\rangle$, ..., $\langle n-1-2s\rangle$ one by one to P_3. The divine rule is violated $n-2-2s$ times.

(b). Now we have $2s+1$ disks left on P_1. Group the top $2s$ into s pairs, then do the same thing as in the stage (2) in the first case. The divine rule is violated s times here.

(c). Move $\langle n\rangle$ to P_2.

(d). For the top s pairs on P_3, do the same thing as in stage (4) in the first case.

(e). The $n-1-2s$ disks are now in the reversed order on P_3. We move them one by one to P_2.

The divine rule is violated $n-2-2s+s=k$ times, and the total number of steps is

$$(n-1-2s)+3s+1+3s+(n-1-2s)=2n+2s-1=4n-2k-5.$$

Finally, if $n \leq k+2$, we simply move the top $n-1$ disks one by one to P_3, and move $\langle n \rangle$ to P_2, then move all the other disks to P_2. □

The rest of this section is devoted to the proof that $S(n,k)=g(n,k)$.

Proposition 5.3.2. *Let $n \geq 1$; let S_1 be the state such that $[n]$ is stacked on one peg in increasing order, and the other two pegs are empty; let S_2 be a state where $[n]$ is on one peg (maybe the same as the original one) in any order, and the other two pegs are empty; let $\mathcal{A}$ be any procedure that takes S_1 to S_2 with at most k violations of the divine rule, and assume $\langle n \rangle$ is moved at least once in $\mathcal{A}$. Then there are at least $2^{n-2k}-1$ steps in $\mathcal{A}$.*

Proof. We prove the proposition by induction on n. The base cases $n=1$ and $n \leq 2k$ are trivial. Now assume $n>1$, $n>2k$, and the theorem is true for all $n'<n$. We color the disk $\langle x \rangle$ black if $\langle x \rangle$ is ever moved onto a smaller disk in $\mathcal{A}$. Otherwise we color $\langle x \rangle$ as white.

Case 1. $\langle n \rangle$ is black. Suppose there are s $(1 \leq s \leq k)$ consecutive black disks at the bottom of the stack in state S_1. They are $\langle n \rangle$, …, $\langle n-s+1 \rangle$. Disk $\langle n-s \rangle$ is white. There are $n-s-1>0$ disks above $\langle n-s \rangle$, we call this set of disks T. Notice that, since $\langle n-s \rangle$ is white, right before the first move of $\langle n-s \rangle$, the whole set T is carried to another peg with at most $k-s$ violations of the divine rule. By induction, the procedure already has at least $2^{n-s-1-2(k-s)}-1=2^{n-2k+s-1}-1 \geq 2^{n-2k}-1$ steps at this moment.

Case 2. $\langle n \rangle$ is white. Focus on the moment right before the first move of $\langle n \rangle$, the whole set $[n-1]$ is moved from one peg to another peg.

Case 2.1. $\mathcal{A}$ violates the divine rule less than k times before this moment, then it has already taken at least $2^{n-1-2(k-1)}-1>2^{n-2k}-1$ steps.

Case 2.2. Otherwise, all the k violations occur before this moment. It has taken at least $2^{n-1-2k}-1$ steps. Consider the state of $[n-1]$ at this moment, call it S'_1. Call the final state of $[n-1]$ S'_2, which is $[n-1]$ in increasing order on a peg, because all the k violations occurred before the state S'_1. Since $\langle n\rangle$ is white, it serves only as ground to $[n-1]$. The procedure $\mathcal{A}$ from this moment, with moves on $\langle n\rangle$ excluded, is a procedure taking S'_1 to S'_2 without any violation of the divine rule. Notice that, in S'_1 there can be at most k disks such that it is bigger than the disk below it. By Lemma 5.2.4 and the inductive hypothesis applied to $\mathcal{A}$, there are at least $2^{n-1-2k}-1$ steps after this moment in $\mathcal{A}$. Counting the moves of $\langle n\rangle$, the total number of steps is at least $2^{n-2k}-1$. □

Proposition 5.3.3. *For any $n \le 2k+2$, $S(n,k) = g(n,k)$.*

Proof. This is trivial when $n \le k+2$. We show $S(n,k) \ge 4n-2k-5$ for all $n \ge k+2$. We color the disk $\langle x\rangle$ black if $\langle x\rangle$ is ever moved onto a smaller disk in $\mathcal{A}$. Otherwise we color $\langle x\rangle$ as white. We assume there are s black disks, $s \le k$. Each black disk is moved at least twice: one violates the divine rule, and the last one does not. So the cost of all the black disks is at least $2s$. Let T be the set $[n-1]$. Consider the state S right before the first move of $\langle n\rangle$.

Case 1. $\langle n\rangle$ is white. All the disks in T are stacked on another peg P. Let $\langle x\rangle$ be any white disk in T. If $\langle x\rangle$ is at the bottom, the cost of $\langle x\rangle$ in the entire procedure is at least 2. Otherwise, $\langle x\rangle$ is on the top of a bigger disk $\langle y\rangle$. $\langle x\rangle$ is moved at least twice before, otherwise $\langle y\rangle$ can not go below $\langle x\rangle$ on P. $\langle x\rangle$ will be moved at least twice later, otherwise $\langle y\rangle$ can not go below $\langle x\rangle$ on the destination peg. So, the cost on the white pegs is at least $4(n-s-2)+2+1$.

Case 2. $\langle n\rangle$ is black. The other disks are stacked on the other two pegs. There are $n-s$ white disks. Similar to the previous case, at most two of them has cost 2, each of the other white disks costs at least 4. So, the cost on the white disks is at least $4(n-s-2)+4$.

In either case, the total cost is at least $4(n-s-2)+3+2s \ge 4n-2k-5$. □

We remark that once we fixed the k black disks, the lower bound holds even in the case where each black disk can violate the divine rule many times. If we want to achieve the lower bound, the bottom disk must be white. However, by choosing different black disks, as long as $\langle 1\rangle$ and $\langle n\rangle$ are white, and there are no adjacent white disks, we can achieve the lower bound in different ways.

While it is possible to provide a more careful induction like the proof of Proposition 5.3.2, we found the following way of proving the main theorem in this section more pleasing.

Theorem 5.3.4. *For any n and k, $S(n,k) = g(n,k)$.*

Proof. The case where $n \leq 2k+2$ is done in Proposition 5.3.3. We prove the theorem for $n \geq 2k+2$ by an induction on n for any fixed k.

The basis is $n = 2k+2$, this is done in Proposition 5.3.3. Now assume $n > 2k+2$. Consider the smallest disk $\langle 1\rangle$. For any procedure, if we ignore all the moves of $\langle 1\rangle$, it is a procedure for $H_S([n]\setminus\{\langle 1\rangle\}, P_1, P_2, k)$. By the inductive hypothesis, the procedure needs at least $g(n-1,k) = 2^{n-1-2k}+6k-1$ steps on $[n]\setminus\{\langle 1\rangle\}$. So, if $\langle 1\rangle$ is ever moved more than $2^{n-1-2k}-1$ steps, the induction is finished.

We prove that the other case, where $\langle 1\rangle$ is moved at most $2^{n-1-2k}-1$ steps, is impossible.

Assuming the contrary, there is a procedure $\mathcal{P}$ for the game $H_S([n], P_1, P_2, k)$ with $c(\mathcal{P}, \langle 1\rangle) \leq 2^{n-1-2k}-1$. Suppose the cost of the procedure $\mathcal{P}$ is N. We pick a large m so that $2^{m+1} > N$. Now, suppose D is a set of m disks smaller than $\langle 1\rangle$. We define a procedure $\mathcal{P}'$ which simulates $\mathcal{P}$. Whenever we have a move $\langle 1\rangle(P \to Q)$, we move the whole tower $\{\langle 1\rangle\} \cup D$ from P to Q in the canonical way with $2^{m+1}-1$ moves. It is clear that the procedure $\mathcal{P}'$ successfully plays the game $H_S([n]\cup D, P_1, P_2, k)$, which is the same as $H_S([n+m], P_1, P_2, k)$. Since $c(\mathcal{P}, \langle 1\rangle) \leq 2^{n-1-2k}-1$, we have $c(\mathcal{P}, \{\langle 1\rangle\}\cup D) \leq (2^{n-1-2k}-1)(2^{m+1}-1)$. So, the total cost of $\mathcal{P}'$ is at most

$$(2^{n-1-2k}-1)(2^{m+1}-1) + N < 2^{m+n-2k} - 2^{m+1} + N \leq 2^{m+n-2k} - 1.$$

Thus contradicts Proposition 5.3.2. ∎

5.4 Santa's towers

Define $C(n, d)$ to be the minimum number of moves one needs to move n disks obeying the Santa's rule; and define $G(n, d)$ to be the minimum number of moves one needs to move n disks obeying the generous Santa's rule. In this section, we derive the exact formula for $C(n, d)$ and $G(n, d)$. Hence, we resolve the open question asked by D. Wood in 1981 ([38]), which asks for a general formula for $G(n, d)$. [1]

It is clear that the Santa's rule with discrepancy d is more restricted than the generous Santa's rule with discrepancy d, hence we have

Fact 5.4.1. $G(n, d) \leq C(n, d)$.

It is also easy to see that the generous Santa's rule is not too generous. There is an implicit restriction on the positions where the smaller disks can occur. We mention the following fact, although we will not need it.

Fact 5.4.2. *For any $d > t > 0$, if $\langle x \rangle$ is placed onto a pile which contains the disk $\langle x - t \rangle$ under the generous Santa's rule with discrepancy d, then $\langle x - t \rangle$ must occur among the top $2d - t - 2$ positions in that pile before the move.*

Proof. Observe that any disk $\langle y \rangle$ above $\langle x - t \rangle$ in that pile must satisfy $y < x - t + d$, $y > x - d$, and $y \neq x$. □

Our main result in this section is to prove $C(n, d) = G(n, d) = 1 + d(2^{q+1} - 2) + r2^{q+1}$. The solution is based on two ways of grouping the disks. To get an upper bound, we group every d consecutive disks, and simulate the canonical `hanoi`, viewing each group as a disk. To prove the lower bound, we view each arithmetic progression with difference d as a class, and analyze the lower bound on each class.

5.4.1 A procedure for $H_C([n], P_a, P_b, d)$

We study the following procedure.

[1] As we reword the problem here, our $G(n, d)$ is equivalent to $f(n, d - 1)$ in [38]. We found some mistakes in the table on page 23 of [38]. As we will show, $f(4, 1) = 9$ and $f(5, 2) = 11$, which are miscalculated as 11 and 13, respectively, in [38].

Procedure 5.4.3. *santa([n], a, b, d)*

Let $n-1 = dq+r$, where $1 \le r \le d$. We group the set $[n-1]$ into q blocks of size d and one block of size r. To be precise, let $B_1 = \{\langle i \rangle : 1 \le i \le r\}$, and $B_j = \{\langle i \rangle : r+(j-2)d+1 \le i \le r+(j-1)d\}$, $2 \le j \le q+1$. We view each block as a disk and simulate the canonical **hanoi** *procedure on the set of blocks. An instruction to move B from P_i to P_j means we move all the disks in block B one by one from P_i to P_j; we call this a block move. Notice that each block move corresponds to up to d actual moves, and the order inside the block is reversed. We denote the set $\{B_i : 1 \le i \le q+1\}$ by $[[q+1]]$.*

```
c := 6-a-b;
hanoi([[q+1]], a, c);
move disk n from peg a to peg b;
hanoi([[q+1]], c, b);
```

Proposition 5.4.4. *The procedure* ***santa([n], a, b, d)*** *successfully plays the game under Santa's rule, $H_C([n], P_a, P_b, d)$.*

Proof. Notice that, when $i < j$, all the disks in block B_i are smaller than the disks in block B_j. We simulate the original Hanoi game on the blocks, never put B_j onto B_i when $i < j$, so the Santa's rule with discrepancy d is obeyed. Finally, the blocks are in the correct order on the peg P_b. We just need to check if the disks are in the right order within each block. This is so because each block move reversed the order within the block, and since we simulated the canonical Hanoi twice, each block is involved in an even number of block moves. □

The following proposition follows easily from Fact 5.2.2.

Proposition 5.4.5. *Let $n-1 = dq+r$, where $1 \le r \le d$. The number of moves that occur in Procedure 5.4.3 is $1 + d(2^{q+1}-2) + r2^{q+1}$.*

5.4.2 The formula for $C(n,d)$ and $G(n,d)$

Theorem 5.4.6. *Let $n-1 = dq + r$, where $1 \le r \le d$;*

$$C(n,d) = G(n,d) = 1 + d(2^{q+1} - 2) + r2^{q+1}.$$

Proof. By Propositions 5.4.4, 5.4.5, and Fact 5.4.1, we have $G(n,d) \le C(n,d) \le 1 + d(2^{q+1} - 2) + r2^{q+1}$. It is enough to prove the lower bound for $G(n,d)$, i.e., Procedure 5.4.3 is the best one can achieve to play the game $H_G([n], P_a, P_b, d)$. We denote Procedure 5.4.3 by $\mathcal{A}$. We partition the disks into d classes. $C_k = \{\langle i \rangle : i \equiv k \pmod d\}$, $0 \le k < d$. Let C_{k_0} be the class to which $\langle n \rangle$ belongs, i.e., $n \equiv k_0 \pmod d$, $0 \le k_0 < d$. We give new names to the disks. For any class C_k with s disks, we label the disks in that class, from the biggest to the smallest, $\langle k \rangle_1, \ldots, \langle k \rangle_s$.

Under the generous Santa's rule, the disks in each class can not reverse their order on any peg. So, if we disregard all the other disks, only observe the movement of disks in C_k, it is a canonical game $H(C_k, P_a, P_b)$. So, the cost involving that class is at least $2^{|C_k|} - 1$ in any successful play of the game $H_G([n], P_a, P_b, d)$. However, it is easy to see, $c(\mathcal{A}, \langle k \rangle_i) = 2^i$ if $k \ne k_0$. $c(\mathcal{A}, \langle k_0 \rangle_i) = 2^{i-1}$. So, whenever $k \ne k_0$, $c(\mathcal{A}, C_k) = 2(2^{|C_k|} - 1)$. Here is a little remark about these values. If our rule does not require that all pegs are in order in the final state, one can easily see that there is a procedure that requires just $2^{|C_k|} - 1$ steps for each C_k. The main difficulty in the proof is to show, in order to make the final order correct, that we really need to double the number of moves for almost all the classes. The parity condition in Lemma 5.4.9(b) is crucial.

From now on, we fix $\mathcal{B}$ to be any procedure that moves $[n]$ from P_1 to P_2 under the generous Santa's rule with discrepancy d.

Definition Two consecutive moves in a procedure are called *redundant* if they are in the form $\langle x \rangle (P \to Q)$ and $\langle x \rangle (Q \to P)$. A procedure is called *reduced* if there are no redundant moves. From any procedure $\mathcal{P}$, we keep deleting redundant moves in any order, it is easy to see we get a unique reduced procedure, we call this the reduced procedure of $\mathcal{P}$, denoted $\rho\mathcal{P}$.

Definition Let $\mathcal{P}$ be a procedure with a sequence of disk moves. Let X be a subset of the disks. The *induced procedure of* $\mathcal{P}$ *with respect to* X, denoted $\mathcal{P} \mid_X$, is the subsequence of $\mathcal{P}$ that involves the moves of disks from X. We call $\rho(\mathcal{P} \mid_X)$ the *reduced procedure of* $\mathcal{P}$ *with respect to* X.

We will use the following easy fact frequently.

Fact 5.4.7. *Let* $\mathcal{P}$ *be any procedure that obeys the generous Santa's rule and moves a disk set* D *from* P_1 *to* P_2. *Let* X *and* Y *be subsets of* D *such that* $X \subseteq Y$. *Then (a)* $\mathcal{P} \mid_X$ *is a procedure that obeys Santa's rule and moves the disk set* X *from* P_1 *to* P_2. *The order of disks in the initial and final states are the same as those in the procedure* $\mathcal{P}$. *The same is true for the procedure* $\rho(\mathcal{P} \mid_X)$. *(b)* $\rho(\mathcal{P} \mid_X)$ *appears as a sub-sequence in* $\rho(\mathcal{P} \mid_Y)$. *Consequently,* $c(\rho(\mathcal{P} \mid_X), \langle x \rangle) \leq c(\rho(\mathcal{P} \mid_Y), \langle x \rangle)$ *for any disk* $\langle x \rangle$. *(c) For any* X *and any disk* $\langle x \rangle$, $c(\rho(\mathcal{P} \mid_X), \langle x \rangle)$ *is of the same parity as* $c(\mathcal{P}, \langle x \rangle)$.

Definition Let $\mathcal{P}$ be a procedure with a sequence of m moves, $(\langle x_i \rangle (P_i \to Q_i) : i = 1 \cdots m)$. The *"log"* of $\mathcal{P}$ is the sequence $(\langle x_i \rangle : i = 1 \cdots m)$.

Lemma 5.4.8. *Let* $\mathcal{P}$ *be any procedure that obeys the generous Santa's rule with discrepancy* d. *Let* X *be any subset of disks such that each disk in* X *is of size at least* $x + d$. *(a) If* $y - x \geq d$ *and* $z - y \geq d$, *then* $\langle y \rangle$ *and* $\langle z \rangle$ *can not both appear between any pair of consecutive* $\langle x \rangle$*'s in the log of* $\mathcal{P}$. *(b) Between any two consecutive* $\langle x \rangle$*-moves in* $\rho(\mathcal{P} \mid_{X \cup \{\langle x \rangle\}})$, *all the moves of disks in* X *are from the same source peg and to the same destination peg. These are the two pegs used,other than the peg containing* $\langle x \rangle$, *during this period. (c) For any* $\langle y \rangle \in X$. *There is at most one* $\langle y \rangle$ *between any pair of consecutive* $\langle x \rangle$*'s in the log of* $\rho(\mathcal{P} \mid_{X \cup \{\langle x \rangle\}})$.

Proof. Let S_1 be the state after an $\langle x \rangle$-move, and let S_2 be the state before the next time we move $\langle x \rangle$. From the state S_1 to S_2, $\langle x \rangle$ stays on the same peg, say, P_1.

(a). If $\langle y \rangle$ is ever moved during the period from S_1 to S_2, then y is never on the peg P_1 during this period. A move for $\langle z \rangle$ is impossible: It would either move above $\langle y \rangle$, or already be above $\langle y \rangle$ before the move.

(b) and (c). During the period between S_1 and S_2, no disk in X can be moved to peg P_1. Because the procedure $\rho(\mathcal{P} \mid_{X\cup\{\langle x\rangle\}})$ is reduced, all the moves during this period have the same source and destination, say, from P_2 to P_3. So, any disk in X will be moved at most once. □

Lemma 5.4.9. *Let $\langle x\rangle$ and $\langle x'\rangle$ be two disks such that $x < x'$. Let X be a set of disks such any disk in X is of size at least $x' + d$. Let $T \subseteq X$ such that the difference of sizes between each pair of disks in T is at least d. Suppose $c(\rho(\mathcal{B} \mid_X), T) \geq s$. Then (a) $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle\}}), \langle x\rangle) \geq s + 1$, $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x'\rangle\}}), \langle x'\rangle) \geq s + 1$. (b) If $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x'\rangle\}}), \langle x'\rangle) = s+1$ and $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}}), \langle x\rangle) = s+1$, then either $s+1$ is an even number, or $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}}), \langle x'\rangle) \geq s + 3$.*

Proof. (a). Look at the procedure $\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle\}})$. Clearly $\langle x\rangle$ is moved before we ever move any disk from T, and is also moved after we finish all the disks from T. Between any two consecutive appearances of symbols from T in the log, we must have at least one appearance of x, otherwise it contradicts Lemma 5.4.8. Since X is a subset of $X \cup \{\langle x\rangle\}$, there are at least s symbols from T in the log, so $c(\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle\}}), \langle x\rangle) \geq s + 1$. The same proof applies to $\langle x'\rangle$ instead of $\langle x\rangle$.

(b). Let $\mathcal{P} = \rho(\mathcal{B} \mid_{X\cup\{\langle x'\rangle\}})$. $c(\mathcal{P}, \langle x'\rangle) \geq s + 1$, and the equality holds only if $c(\mathcal{P}, T) = s$, and the $s + 1$ $\langle x'\rangle$'s and the s symbols from T appear alternatively in the log of $\mathcal{P}$. Let the i-th move from T be $\langle y_i\rangle(R_i \to Q_i)$, and let the i-th move of $\langle x'\rangle$ be $\langle x'\rangle(R'_i \to Q'_i)$. It is easy to see that $\{R_i, Q_i\} = \{P_1, P_2, P_3\} \setminus \{R'_{i+1}\}$, and $\{R_{i+1}, Q_{i+1}\} = \{P_1, P_2, P_3\} \setminus \{Q'_{i+1}\}$. Since $R'_{i+1} \neq Q'_{i+1}$, we conclude $\{R_i, Q_i, R_{i+1}, Q_{i+1}\} = \{P_1, P_2, P_3\}$.

The procedure $\mathcal{P}$ is a sub-sequence of $\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$. Focus on the s moves $\langle y_i\rangle(R_i \to Q_i)$ in the procedure $\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$. We call them the T-moves. The $s + 1$ x's must appear once before all the T-moves, once after all the T-moves, and once between each consecutive T-moves. The last statement is true, otherwise there are two consecutive T-moves $\langle y_i\rangle(R_i \to Q_i)$ and $\langle y_{i+1}\rangle(R_{i+1} \to Q_{i+1})$ while $\langle x\rangle$ stays on a certain peg, which is impossible since $\{R_i, Q_i, R_{i+1}, Q_{i+1}\} = \{P_1, P_2, P_3\}$.

In particular, we proved that no two $\langle x\rangle$-moves in $\rho(\mathcal{B} \mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$ are adjacent.

This implies that if we delete all the $\langle x\rangle$-moves from $\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$, it is either already reduced, i.e., it is $\rho(\mathcal{B}\mid_{X\cup\{\langle x'\rangle\}})$, or at least two $\langle x'\rangle$-moves are redundant. To see this, observe that the only way to start a reduction is that there is a move, say $\langle x\rangle(P_1 \to P_2)$, where the previous move, which is from $X \cup \{\langle x'\rangle\}$, and the next move reverse each other. However, if these are not $\langle x'\rangle$-moves, the previous move is between P_2 and P_3, and the next move is between P_1 and P_3, since each disk in X is of size at least $x + d$.

As a consequence, if $c(\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}}), \langle x'\rangle) < s + 3$, then, because

$$c(\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}}), \langle x'\rangle) \geq c(\rho(\mathcal{B}\mid_{X\cup\{\langle x'\rangle\}}), \langle x'\rangle) = s + 1$$

and they have the same parity, there are exactly $s+1$ $\langle x'\rangle$-moves and exactly s T-moves in the procedure $\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$. The $\langle x\rangle$-moves appear once before all the moves from T, once after all of them, and once between each consecutive moves from T. The same is true for x'. We further claim that the source and destination of each move is fully determined by the moves of elements in T. We prove this claim inductively from the first move to the last one. Let $\langle x\rangle(P \to Q)$ be the i-th move of $\langle x\rangle$. If it is the first one, then $P = P_1$, otherwise it is the destination of the last move. If it is the last move, $Q = P_2$. Otherwise, the next move from T is $\langle y_i\rangle(R_i \to Q_i)$, Q must be the peg other than R_i and Q_i.

Thus, we have $s + 1$ positions in the log of $\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle,\langle x'\rangle\}})$, one before all the symbols from T, one after them, and one between each pair of consecutive symbols from T. In each position, there is one $\langle x\rangle$ and one $\langle x'\rangle$, with the moves from the same source to the same destination. Assume $x < x'$, in the odd numbered positions x moves first, and in the even numbered positions x' moves first. If $s + 1$ is odd, x would be stacked under x' in the final state, a contradiction. □

Lemma 5.4.10. *Let X be a set of disks such that each disk in X is of size at least $x + d$. Let $T \subseteq X$ be a subset of size at least 2 such that the difference of sizes between each pair of disks in T is at least d, except for one pair $\langle y\rangle$ and $\langle y'\rangle$. Suppose $c(\rho(\mathcal{B}\mid_X), T \setminus \{\langle y\rangle, \langle y'\rangle\}) \geq s$, $c(\rho(\mathcal{B}\mid_X), \langle y\rangle) \geq p$, and $c(\rho(\mathcal{B}\mid_X), \langle y'\rangle) \geq p - 1$ and has a different parity than p. Then $c(\rho(\mathcal{B}\mid_{X\cup\{\langle x\rangle\}}), x) \geq s + p + 2$.*

Proof. By Lemma 5.4.9 with the set $T \setminus \{\langle y' \rangle\}$, x appears at least $s+p+1$ times in $\rho(\mathcal{B} \mid_{X \cup \{x\}})$. Assuming the contrary, x appears exactly $s+p+1$ times. Then there are exactly $s+p$ symbols from $T \setminus \{\langle y' \rangle\}$ in the log of $\rho(\mathcal{B} \mid_{X \cup \{x\}})$. Consider the $s+p$ spots, one between each consecutive pair of x's in the log. By Lemma 5.4.8, the $s+p$ symbols from $T \setminus \{\langle y' \rangle\}$ appears once in each spot, and $\langle y' \rangle$ can not be put in a spot twice or in the same spot with any disk other than $\langle y \rangle$. Since there are at most p such spots, and $c(\rho(\mathcal{B} \mid_{X \cup \{\langle x \rangle\}}), x)$ is not the same parity as p, we must have exactly $p-1$ $\langle y' \rangle$-moves. We call a spot even if it contains both $\langle y \rangle$ and $\langle y' \rangle$, otherwise call it odd. Since neither $\langle y \rangle$ nor $\langle y' \rangle$ can be moved onto the peg containing $\langle x \rangle$, we must have the same source and destination for $\langle y \rangle$ and $\langle y' \rangle$ in each even spot. Consider the only odd spot. If it is the last move of $\langle y \rangle$, $\langle y \rangle$ and $\langle y' \rangle$ would be on different pegs in the final state. Otherwise $\langle y \rangle$ and $\langle y' \rangle$ can not move from the same source in the next even spot. We have a contradiction in either case. □

Definition For any disk $\langle x \rangle$, define $U(\langle x \rangle)$ to be the set of disks that are not smaller than x, i.e., $U(\langle x \rangle) = \{\langle x \rangle, \langle x+1 \rangle, \ldots, \langle n \rangle\}$. Define the *downwards cost* of $\langle x \rangle$ to be $d(\langle x \rangle) = c(\rho(\mathcal{B} \mid_{U(\langle x \rangle)}), \langle x \rangle)$, i.e., the number of moves of $\langle x \rangle$ in the reduced procedure of $\mathcal{B}$ if we disregard all the smaller disks.

Let C_k be any class where $k \neq k_0$. It is easy to see that $c(\mathcal{A}, C_k) = 2(2^{|C_k|} - 1)$. If $c(\mathcal{B}, C_k) < 2(2^{|C_k|} - 1)$, there must be a disk $\langle k \rangle_i$ such that $c(\mathcal{B}, \langle k \rangle_i) < 2^i$; so, the disk $\langle k \rangle_i$ satisfies $d(\langle k \rangle_i) < 2^i$. Thus, we have the following definition.

Definition We call a class C_k *bad* if $k \neq k_0$ and $c(\mathcal{B}, C_k) < 2(2^{|C_k|} - 1)$. Let C_k be a bad class, there must be a disk $\langle k \rangle_i$ such that $d(\langle k \rangle_i) < 2^i$. We call a disk $\langle k \rangle_i$ *precious* if $d(\langle k \rangle_i) < 2^i$. We call the biggest precious disk in C_k the *prince* of C_k.

Lemma 5.4.11. *(a) Let C_k be a bad class with prince $\langle k \rangle_x$. Then $\langle k \rangle_x$ is the only precious disk in C_k, and $c(\mathcal{B}, \langle k \rangle_x) = 2^x - 1$. (b) Suppose C_{k_1} and C_{k_2} are two bad classes with princes $\langle k_1 \rangle_y$ and $\langle k_2 \rangle_z$, respectively. Then $y \neq z$.*

Proof. (a). By the definition of a prince, $d(\langle k \rangle_i) \geq 2^i$ for each $i < x$. Set $T = \{\langle k \rangle_i : i < x\}$ and $X = U(\langle k \rangle_{x-1})$. We have $c(\rho(\mathcal{B} \mid_X), \langle k \rangle_i) \geq d(\langle k \rangle_i) \geq 2^i$. By Lemma 5.4.9(a)

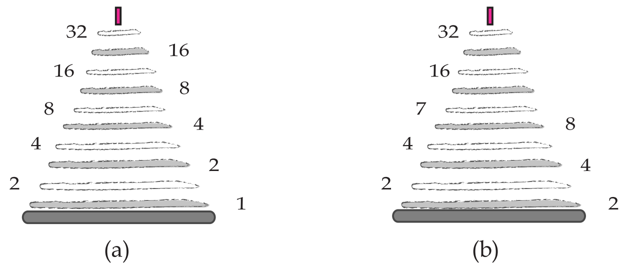

Figure 5.2: The gray disks are the class C_{k_0}, the white disks are the class C_k. (a) The cost of each disk in Procedure 5.4.3. (b) The least possible costs if C_k is a bad class with its third disk as its prince.

we get $d(\langle k\rangle_x) \geq c(\rho(\mathcal{B}\mid_{X\cup\{\langle k\rangle_x\}}), \langle k\rangle_x) \geq 2^x - 1$. In order to make $\langle k\rangle_x$ a prince, we must have $d(\langle k\rangle_x) = 2^x - 1$ and $d(\langle k\rangle_i) = 2^i$ for each $i < x$.

We claim that one of two things must happen:

$$d(\langle k_0\rangle_i) \geq 2^i, \forall i \leq x; \text{ or} \tag{5.1}$$

$$d(\langle k_0\rangle_i) \geq 2^i, \forall i < x, \text{ and } c(\rho(\mathcal{B}\mid_{U(\langle k\rangle_x)}), \langle k_0\rangle_x) \geq 2^x + 1. \tag{5.2}$$

To prove the claim, assume there is $i \leq x$ such that $d(\langle k_0\rangle_i) < 2^i$ and $d(\langle k_0\rangle_j) \geq 2^j$ for all $j < i$. Setting $T = \{\langle k_0\rangle_j : j < i\}$ and $X = U(\langle k_0\rangle_{i-1})$ in Lemma 5.4.9, we have $d(\langle k_0\rangle_i) = 2^i - 1$ and $d(\langle k_0\rangle_j) = 2^j$ for all $j < i$. (The case $i = 1$ is trivial.) We discuss two other cases. In the first case, $i = x$. Consider the set $T = \{\langle k_0\rangle_j : j < x\}$ and $X = U(\langle k_0\rangle_{x-1})$. $c(\rho(\mathcal{B}\mid_X), T) \geq 2^x - 2$. By Lemma 5.4.9(a), $c(\rho(\mathcal{B}\mid_{X\cup\{\langle k_0\rangle_x\}}), \langle k_0\rangle_x) \geq 2^x - 1$. On the other hand, we have

$$c(\rho(\mathcal{B}\mid_{X\cup\{\langle k_0\rangle_x\}}), \langle k_0\rangle_x) \leq c(\rho(\mathcal{B}\mid_{U(\langle k_0\rangle_x}), \langle k_0\rangle_x) = d(\langle k_0\rangle_x) = 2^x - 1.$$

So, $c(\rho(\mathcal{B}\mid_{X\cup\{\langle k_0\rangle_x\}}), \langle k_0\rangle_x) = 2^x - 1$. Similarly, we have $c(\rho(\mathcal{B}\mid_{X\cup\{\langle k_0\rangle_x, \langle k\rangle_x\}}), \langle k\rangle_x) = 2^x - 1$. By Lemma 5.4.9 (b) we establish (5.2). In the second case where $i < x$, using Lemma 5.4.10 with $T = \{\langle k_0\rangle_j : j \leq i\} \cup \{\langle k\rangle_j : i \leq j < x\}$, $X = U(\langle k\rangle_{x-1})$, $\langle y\rangle = \langle k\rangle_i$, and $\langle y'\rangle = \langle k_0\rangle_i$, we get the cost $d(\langle k\rangle_x) \geq 2^x$. It is not a prince.

Having proved the claim, we now show that $d(\langle k\rangle_i) \geq 2^i$ for any $i > x$. We do this inductively on i. If there is an $t \leq x$ such that $c(\rho(\mathcal{B}\mid_{U(\langle k\rangle_x)}), \langle k_0\rangle_t) \geq 2^t + 1$, we may set $T_i = \{\langle k_0\rangle_j : j \leq x\} \cup \{\langle k\rangle_j : x < j < i\}$ for each i and use Lemma 5.4.9(a). Otherwise, by (5.1) and (5.2), $c(\rho(\mathcal{B}\mid_{U(\langle k\rangle_x)}), \langle k_0\rangle_t) = 2^t$ for each $t \leq x$. Set $T_i = \{\langle k_0\rangle_j : j \leq x\} \cup \{\langle k\rangle_j : x \leq j < i\}$, $X_i = U(\langle k\rangle_{i-1})$ and use Lemma 5.4.10. In both cases, the cost for $\langle k\rangle_j$ is at least 2^j for all $j > x$. By the inductive hypothesis, we get $c(\mathcal{B}, \langle k\rangle_i) \geq 2^i$.

Finally, we must have $c(\mathcal{B}, \langle k\rangle_x) = 2^x - 1$, otherwise C_k is not a bad class.

(b). We may assume the biggest disk in C_{k_1} is bigger than the biggest disk in C_{k_2}, so $\langle k_1\rangle_i$ is bigger than $\langle k_2\rangle_i$ for any i. Assuming the contrary $y = z$, consider the set $T = \{\langle k_1\rangle_i : i < y\}$ and $X = U(\langle k_1\rangle_{y-1})$. By Lemma 5.4.9 with $\langle k_1\rangle_y$ and $\langle k_2\rangle_y$, $c(\mathcal{B}, \langle k_1\rangle_y) \geq 2^y + 1$, a contradiction. □

Lemma 5.4.12. *If there is any bad class, then* $c(\mathcal{B}, C_{k_0}) \geq c(\mathcal{A}, C_{k_0}) + 2^{|C_{k_0}|-1}$.

Proof. By the discussion in the proof of Lemma 5.4.11(a), if there is any bad class, $c(\mathcal{B}, \langle k_0\rangle_1) \geq 2$. In the procedure $\mathcal{B}$, before the first move of $\langle k_0\rangle_1$, after the last of its moves, and between each consecutive pair of its moves, the set $C_{k_0} \setminus \{\langle k_0\rangle_1\}$ is moved completely from a peg to another under the divine rule at least three times. Notice that the same set is moved twice under the divine rule in $\mathcal{A}$. So, the extra cost on this set is at least $2^{|C_{k_0}|-1} - 1$. Adding the extra cost on $\langle k_0\rangle_1$, the lemma follows. □

Now we finish our proof of the main theorem. We compare the cost of the procedures $\mathcal{A}$ and $\mathcal{B}$ on each individual disk. By Lemma 5.4.11 (a), $\mathcal{B}$ can only save one step against $\mathcal{A}$ for each prince. However, if there is any bad class, $\mathcal{A}$ saves at least $2^{|C_{k_0}|-1} \geq |C_{k_0}|$ steps against $\mathcal{B}$ for the disks in the class C_{k_0} by Lemma 5.4.12. By Lemma 5.4.11 (b), the number of princes is no more than $|C_{k_0}|$. In fact, $\mathcal{B}$ can not save as many as $|C_{k_0}|$ steps on the princes, since a prince in the range $[n-d, n]$ still needs to moved at least twice. □

5.5 Other towers of hanoi

There could be many reasonable variations of the Hanoi tower problem. We mention one here.

As the remark after the proof of Proposition 5.3.3 shows, when $n \leq 2k+2$, one needs $g(n,k)$ steps even when we fix k disks and each of them can violate the divine rule many times. We propose the following variation of our Sinner's mode.

> *Hanoi Tower with k Evildoers*: Among the n disks, one may pick any k of them as evildoers then perform a procedure of disk moves. An ordinary disk can not be placed directly on the top of a smaller disk, but an evildoer can do this an unlimited number of times.

Let $E(n,k)$ be the minimum number of steps one need to move the entire stack from one peg to another in the evildoer's mode, where the disks are in increasing order in both initial and final configuration. Clearly $E(n,k) \leq S(n,k)$. A computer verification shows that $E(n,k) = S(n,k)$ for all $n < 8$ and any k. However, the program shows that $E(8,1) = 57$ while $S(8,1) = 69$. It could be an interesting problem to find the formula for $E(n,k)$.

We describe the procedure in 57 moves that transforms the 8 disks from P_1 to P_3, with the evildoer $\langle 6 \rangle$.

- Steps 1 to 15: Move the 4 smallest disks from P_1 to P_3.
- Step 16: $\langle 5 \rangle (P_1 \to P_2)$.
- Step 17: $\langle 6 \rangle (P_1 \to P_3)$ — notice that $\langle 6 \rangle$ is an evildoer.
- Steps 18 to 21: $\langle 5 \rangle (P_2 \to P_3)$, $\langle 7 \rangle (P_1 \to P_2)$, $\langle 5 \rangle (P_3 \to P_2)$, then $\langle 6 \rangle (P_3 \to P_1)$.
- Steps 22 to 24: Move $\langle 1 \rangle$ and $\langle 2 \rangle$ to P_2.
- Step 25: $\langle 6 \rangle (P_1 \to P_2)$ — does the evil again.
- Steps 26 to 28: Move $\langle 3 \rangle$ and $\langle 4 \rangle$ to P_2.
- Step 29: $\langle 8 \rangle (P_1 \to P_3)$.
- Steps 30 to 57 are symmetric to the first 28 steps.

A computer exhaustive search shows that this is an optimal solution for 8 disks.

Chapter 6

Conclusions and Further Research

In chapter 2, we discussed the most balanced partition of convex bodies in the plane by three non-concurrent lines. We showed that triangle is the extreme convex body that achieves greatest (min area)/(MAX area) ratio at $\frac{2}{1+2\sqrt{2}} \approx 0.522$. Any convex body which is not a triangle has an optimal ratio smaller than that. Also, we defined balanced partitions, a special class of partitions, which are deeply related with this problem. We showed that for any convex body K and a vector v, there are exactly two balanced partitions of K with one of the three lines having normal vector v. We conjectured that any most balanced partition of convex body K must be a balanced partition of K, while we can not prove that.

Our research on balanced partition raised the following open question: Let K be a convex body in the plane and $l_i, \cdots, l_j$ be j lines in general positions and meeting only inside K. $l_i, \cdots, l_j$ partition K into $(j^2+j)/2+1$ regions. If we define $f_K(l_i, \cdots, l_j)$ to be the number of distinct areas of $(j^2+j)/2+1$ regions, what is $\psi_K(j) = min_{\{l_i,\cdots,l_j\}} f_K(l_i, \cdots, l_j)$? By our existence proof of balanced partitions, we have $\psi_K(3) = 2$ for all K.

The zigzag problem is studied in chapter 3. We improved the upper bound to $\frac{n}{2} + O\left(\frac{n}{\log n}\right)$ if self-intersection is allowed and to $n - \lfloor\frac{n-2}{8}\rfloor$ if self-intersection is not allowed. The lower bound seems to be extremely hard to tackle. We conjectured that for big n, there always exists a point set P of size n, which requires at least $n/2+\infty$ turns. Though we are even unable to construct point sets which forces $n/2+2$ turns for big n.

In chapter 4, We have defined and studied forward degree sequences and their associated polynomials. In particular, the properties of (strongly) balanced forward degree sequences were investigated. Our proof shows that any chordal graph has a strongly balanced forward degree sequence and any graph with all degrees at most 3 has a balanced forward degree sequence. Moreover, these (strongly) balanced forward degree sequences can be computed in polynomial time. Our results might bring a new clue for graph problems related to vertex ordering, such as graph isomorphism, because a (strongly) balanced forward degree sequence is an optimal vertex ordering. Our work could be extended by finding more classes of graphs which are (strongly) balanced. Also, it is still open that whether some of the graph invariants derived naturally from forward degree sequence, such as $N(G)$, are polynomial time computable.

In the last chapter, we studied two new variants of the Towers of Hanoi problem. In both variations, one is allowed to put a bigger disk directly on the top of a smaller one under some restrictions. We give procedures to solve these two versions, and prove the optimality of our procedures. Our solution also resolves a problem, which is similar to one of our versions, proposed by D. Wood tweenty-four years ago.

References

[1] I. Bárány, A. Hubard and J. Jerónimo "Slicing Convex Sets and Measures by a Hyperplane." *Discrete and Computational Geometry 39 (1-3)*, 67-75, (2008).

[2] R. Buck and E. Buck. "Equipartitioning of Convex Sets." *Math. Mag. 22 (4)*, 195-198, (1949).

[3] R. Courant and H. Robbins. "What is Mathematics?" *Oxford University Press*, (1941).

[4] M. Sholander. "Proof of a Conjecture of R.C. Buck and E.F. Buck." *Math. Mag. 24 (1)*, 7-19, (1950).

[5] N. Megiddo. "Partitioning with two lines in the plane." *J. Algorithms*, 430-433, (1985).

[6] William Steiger, Mario Szegedy and Jihui Zhao "Six-way equipartitioning by three lines in the plane." *Proceedings of Canadian Conference on Computational Geometry*, 277–280, (2010).

[7] William Steiger, Mario Szegedy, Lei Wang and Bahman Kalantari "Equitable Partitioning with Three Lines." *21st Fall Workshop on Computational Geometry 2011*, November 4-5, (2011).

[8] D. Avis. "On the Partitionability of Point Sets in Space." *Proc. First ACM Symposium on Computational Geometry*, 116-120, (1985).

[9] Imre Bárány and Jirí Matousek. "Equipartition of Two Measures by a 4-Fan." *Discrete and Computational Geometry 27*, 293-301, (2002).

[10] Imre Bárány and Jirí Matousek. "Simultaneous Partitions of Measures by k-fans." *Discrete and Computational Geometry 25 (3)*, 317-334, (2001).

[11] S. Bereg. "Equipartitions of Measures by 2-Fans." *Discrete and Computational Geometry 34 (1)*, 87-96, (2005).

[12] S. Bespamyatnikh, D. Kirkpatrick, and J. Snoeyink. "Generalizing Ham-Sandwich Cuts to Equitable Subdivisions." *Discrete and Computational Geometry 24*, 605-622, (2000).

[13] T. Sakai. "Balanced Convex Partitions of Measures in $\mathbb{R}^2$." *Graphs and Combinatorics 18*, 169-192, (2002).

[14] L. Schulman. "An Equipartition of Planar Sets." *Discrete and Computational Geometry 9*, 257-266, (1992).

[15] W. Steiger and J. Zhao. "Generalized Ham-Sandwich Cuts for Well-Separated point Sets." *Proceedings of the* 20^{th} *Canadian Conference on Computational Geometry, Montreal*, August 2008.

[16] W. Steiger and J. Zhao. "Generalized Ham-Sandwich Cuts." *Discrete and Computational Geometry 44*, 535-545, (2010).

[17] Kranakis, Krizanc and Meertens. "Link length of rectilinear Hamiltonian tours in grids." *Ars Combinatoria 38*, 177-192, (1994).

[18] Bereg, Sergey and Bose, Prosenjit and Dumitrescu, Adrian and Hurtado, Ferran and Valtr, Pavel. "Traversing a set of points with a minimum number of turns." *Proceedings of the twenty-third annual symposium on Computational geometry, SCG '07*, 46-55, (2007).

[19] B. Bollobás. "Modern graph theory." *Springer-Verlag,* New York, (1998).

[20] P. Erdos and T. Gallai. "Graphs with Prescribed Degrees of Vertices." *Mat. Lapok. 11,* 264-274, (1960).

[21] S. Hakimi. "On the Realizability of a Set of Integers as Degrees of the Vertices of a Graph." *SIAM J. Appl. Math. 10,* 496-506, (1962).

[22] V. Havel. "A Remark on the Existence of Finite Graphs." *Casopis Pest. Mat. 80,* 477-480, (1955).

[23] C. M. Hoffmann. "Group-theoretic algorithms and graph isomorphism." *volume 136 of Lecture Notes in Computer Science,* Springer-Verlag Inc., New York, (1982).

[24] E. M. Luks. "Isomorphism of bounded valence can be tested in polynomial time." *Proc. of the 21st Annual Symposium on Foundations of Computing,* 42-49. IEEE, (1980).

[25] D. B. West. "Introduction to Graph Theory." *Prentice Hall,* New Jersey (1996).

[26] M. D. Atkinson. "The cyclic towers of Hanoi." *Inform. Process. Lett. 13 (3),* 118-119, (1981).

[27] X. Chen and J. Shen. "On the Frame-Stewart conjecture about the Towers of Hanoi." *SIAM J. Comput. 33 (3),* 584-589, (2004).

[28] N. Claus (= E. Lucas). "La Tour d Hanoi." *Jeu de calcul, Science et Nature 1 (8),* 127-128, (1884).

[29] Y. Dinitz and S. Solomon. "Optimality of an Algorithm Solving the Bottleneck Tower of Hanoi Problem." *a manucript.*

[30] J. S. Frame. "Solution to advanced problem 3918." *Amer. Math. Monthly 48,* 216-217, (1941).

[31] S. Klavžar, U. Milutinović, and C. Petr. "On the Frame-Stewart algorithm for the multi-peg Tower of Hanoi problem." *Discrete Appl. Math. 120 (1-3),* 141-157, (2002).

[32] C. S. Klein and S. Minsker. "The super Towers of Hanoi problem: large rings on small rings." *Discrete Math. 114 (1-3),* 283-295, (1993).

[33] D. Poole. "The Bottleneck Towers of Hanoi problem." J. *Recreational Math. 24 (3),* 203-207, (1992).

[34] B. M. Stewart. "Advanced problem 3918." *Amer. Math. Monthly 46,* 363, (1939).

[35] B. M. Stewart. "Solution to advanced problem 3918." *Amer. Math. Monthly 48,* 217-219, (1941).

[36] P. K. Stockmeyer. "The Tower of Hanoi: A Historical Survey and Bibliography." *manuscript available at* `http://www.cs.wm.edu/~pkstoc/biblio.ps`.

[37] M. Szegedy. "In how many steps the k peg version of the Towers of Hanoi game can be solved?" *STACS 99 (Trier), Lecture Notes in Comput. Sci., 1563,* Springer, Berlin, 356-361, (1999).

[38] D. Wood, "Towers of Brahma and Hanoi Revisited." *J. Recr. Math. 14,* 17-24, (1981).

Vita

Lei Wang

2012 Ph. D. in Computer Science, Rutgers University

1998-2001 M. Sc. in Computer Science from Shanghai Jiaotong University

1995-1998 B. Sc. in Computer Science from Shanghai Jiaotong University