Random Combinatorial Structures and Randomized Search Heuristics

Dissertation

zur Erlangung des Grades des
Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultäten
der Universität des Saarlandes

vorgelegt von

Daniel Johannsen

Saarbrücken
7. April 2010

Tag des Kolloquiums:
15. Juli 2010

Dekan der Naturwissenschaftlich-Technischen Fakultät I:
Prof. Dr. Holger Hermanns

Prüfungsausschuss:
Prof. Dr. Benjamin Doerr
Prof. Dr. Kurt Mehlhorn
Prof. Dr. Thomas Lengauer (Vorsitzender des Prüfungsausschusses)
Dr. Tobias Friedrich (Akademischer Beisitzer)

Berichterstatter:
Prof. Dr. Benjamin Doerr
Prof. Dr. Kurt Mehlhorn
Prof. Xin Yao

Abstract

This thesis is concerned with the probabilistic analysis of random combinatorial structures and the runtime analysis of randomized search heuristics.

On the subject of random structures, we investigate two classes of combinatorial objects. The first is the class of planar maps and the second is the class of generalized parking functions. We identify typical properties of these structures and show strong concentration results on the probabilities that these properties hold. To this end, we develop and apply techniques based on exact enumeration by generating functions. For several types of random planar maps, this culminates in concentration results for the degree sequence. For parking functions, we determine the distribution of the *defect*, the most characteristic parameter.

On the subject of randomized search heuristics, we present, improve, and unify different probabilistic methods and their applications. In this, special focus is given to potential functions and the analysis of the drift of stochastic processes. We apply these techniques to investigate the runtimes of evolutionary algorithms. In particular, we show for several classical problems in combinatorial optimization how drift analysis can be used in a uniform way to give bounds on the expected runtimes of evolutionary algorithms.

Zusammenfassung

Diese Dissertationsschrift beschäftigt sich mit der wahrscheinlichkeitstheoretischen Analyse von zufälligen kombinatorischen Strukturen und der Laufzeitanalyse randomisierter Suchheuristiken.

Im Bereich der zufälligen Strukturen untersuchen wir zwei Klassen kombinatorischer Objekte. Dies sind zum einen die Klasse aller kombinatorischen Einbettungen planarer Graphen und zum anderen eine Klasse diskreter Funktionen mit bestimmten kombinatorischen Restriktionen (*generalized parking functions*). Für das Studium dieser Klassen entwickeln und verwenden wir zählkombinatorische Methoden die auf erzeugenden Funktionen basieren. Dies erlaubt uns, Konzentrationsresultate für die Gradsequenzen verschiedener Typen zufälliger kombinatorischer Einbettungen planarer Graphen zu erzielen. Darüber hinaus erhalten wir Konzentrationsresultate für den charakteristischen Parameter, den *Defekt*, zufälliger Instanzen der untersuchten diskreten Funktionen.

Im Bereich der randomisierten Suchheuristiken präsentieren und erweitern wir verschiedene wahrscheinlichkeitstheoretische Methoden der Analyse. Ein besonderer Fokus liegt dabei auf der Analyse der Drift stochastischer Prozesse. Wir wenden diese Methoden in der Laufzeitanalyse evolutionärer Algorithmen an. Insbesondere zeigen wir, wie mit Hilfe von Driftanalyse die erwarteten Laufzeiten evolutionärer Algorithmen auf verschiedenen klassischen Problemen der kombinatorischen Optimierung auf einheitliche Weise abgeschätzt werden können.

Diese Arbeit ist in englischer Sprache verfasst.

Acknowledgments

I want to thank my supervisor Benjamin Doerr. I am grateful for his great support during my PhD studies. His constant guidance was of immense help while at the same time he granted me the freedom to find and pursue my own line of research. He was a permanent source of inspiration. I also want to thank Kurt Mehlhorn for providing me with the opportunity to work in the unique research environment of his group at the Max Planck Institute for Informatics.

Many thanks to all my coauthors, especially to Benjamin Doerr, Peter Cameron, Thomas Prellberg, Konstantinos Panagiotou, and Pascal Schweitzer. The results of our collaboration contributed strongly to the contents of this thesis. I am particularly indebted to the latter two who were always there for fruitful discussions and for feedback on earlier versions of this thesis.

Contents

1 Introduction

General purpose algorithms are solution strategies that can be applied to a wide range of problems. Examples of such algorithms are linear programming, greedy algorithms, dynamic programming, branch and bound algorithms, and randomized search heuristics (see, for example, Mehlhorn and Sanders (2009)). Such generic algorithms can be easily implemented, adapted, and reused. They are frequently applied when no or little knowledge on a given problem is available. As a drawback, it is hard to tell for which problems it is useful to apply them. To this end, we address this subject from two perspectives.

The first focusses on a particular problem class and asks

"Which general purpose algorithms successfully solve this problem?"

The second focusses on a particular generic algorithmic approach and asks

"Which problems does this general purpose algorithm solve successfully?"

We address these two questions by two orthogonal approaches. To answer the first question, we focus on the actual instances of specific problems. We take an adversarial view and ask for properties of these instances that cause the problem to be tractable or hard. Moreover, we ask whether these properties are *typical*, that is, occur with high probability in a random problem instance. This leads us to the first subject of this thesis, the analysis of the typical properties of random combinatorial structures. We apply exact counting techniques and probabilistic methods to approach this subject.

The second question focusses on the properties of a specific algorithm. A central property of an algorithm is its runtime. Therefore, our approach to the second question is runtime analysis. We are particularly interested in randomized search heuristics. For these, the runtime is not deterministic but described by a random variable. The runtime analysis of randomized search heuristics is the second subject of this thesis. The techniques we apply have a strong emphasis on drift analysis and are again probabilistic.

1.1. Random Combinatorial Structures

The first part of this thesis addresses our first question. For the class of NP-hard problems, no algorithms with polynomial runtime guarantees are known. However, instances of NP-hard problems that appear in practical applications can often be solved routinely and effectively by general purpose approaches like those mentioned above.

One reason for this is that contrary to a worst case instance, a typical instance of such a problem may have certain properties that can be exploited by a heuristic algorithm. Therefore, we devote the first part of the thesis to the typical properties of random combinatorial structures.

The structures we study are several types of planar maps, on the one hand, and generalized parking functions, on the other. We discuss each type of structure in a self-contained chapter, planar maps in Chapter 2 and parking functions in Chapter 3.

1.1.1. Planar Maps

Typical instances of graphs have been studied extensively in the $\mathcal{G}_{n,p}$ random graph model, where the edges of a graph of size n are present mutually independently with probability p. However, real-world graph instances often follow certain restriction or belong to specific graph classes.

This leads us to the investigation of random graphs from constrained classes, for example, random planar graphs. Here, the challenge lies in the strong stochastic dependencies imposed by the constraints. For example, the existence of certain edges in a random planar graph may exclude that of others. Already the task to formulate an efficient algorithm to uniformly generate a random planar graph of given size is far from trivial, see for example Bodirsky, Gröpl, Johannsen, and Kang (2007).

A seminal result in this area (Duchon, Flajolet, Louchard, and Schaeffer (2004)) was the introduction of *Boltzmann samplers*. These algorithms rely on the exact enumeration of the respective structures by means of generating functions. They generate random graphs such that the resulting distribution satisfies three key properties. First, all graphs of the same size are equally likely. Second, by carefully choosing the algorithm's main parameter, it is possible to enforce that graphs of size n are drawn with a probability in the order of $n^{-\alpha}$. Third, these random graphs decompose into subgraphs which are stochastically independent. The first two properties imply the possibility to uniformly generate a random graph of given size in an expected time of order n^{α} by using the technique of rejection sampling.

Only recently, the concept of Boltzmann samplers also gave rise to structural insights into uniformly generated random graphs (Bernasconi, Panagiotou, and Steger (2008)). The key idea is to conceptually assume that the random object in question was generated by a Boltzmann sampler. Tracing back the steps of the algorithm then allows to exploit the third property described above, the stochastic independence of specific substructures.

In Chapter 2, we extend this technique and apply it to analyze the vertex degree distribution in random (planar) maps. *Maps* are planar graphs embedded in the plane, and are commonly used to describe the topology of geometric arrangements.

In particular, the class of all 3-connected planar maps is combinatorially equivalent to the class of all 3-dimensional convex polyhedra.

Since the pioneering work of Tutte (1963), maps have become popular combinatorial objects, and in the meantime a rich theory studying various aspects of maps has evolved (see Chapter 2). However, little is known about statistical properties of random maps, that is, maps drawn uniformly at random from the class of all maps with a given number of edges. A major achievement in this context is the precise description of the so-called *core-size* of a random map, which was provided by Banderier, Flajolet, Schaeffer, and Soria (2001). Among several other results, they showed that a random map contains typically a giant (that is, linear-size) biconnected submap.

In Chapter 2, we study the degree sequences of several types of random maps, which, among others, includes fundamental map classes like those of biconnected maps, 3-connected maps, and triangulations. In particular, we develop a general framework that allows us to derive relations and exact asymptotic expressions for the expected number of vertices of degree k in random maps from these classes, and also provide accompanying large deviation statements. More precisely, building on the work of Gao and Wormald (2003) on random general maps, we obtain as results of our framework precise information about the number of vertices of degree k in random biconnected, 3-connected, loopless, and bridgeless maps.

Indication of source. The content of Chapter 2 has been previously published in the *Proceedings of SODA '10* (Johannsen and Panagiotou (2010)).

1.1.2. Parking Functions

In Chapter 3, we consider discrete functions of the type $f\colon [m] \to [n]$, where $[k]$ denotes the set $\{1, \ldots, k\}$ for all $k \in \mathbb{N}$. A *parking function* is a function $f\colon [n] \to [n]$ such that for every $k \in [n]$, the preimage of the set $[k]$ is of size at least k, that is, $|f^{-1}\big([k]\big)| \geq k$.

In the context of hashing, Konheim and Weiss (1966) showed that there exists exactly $(n+1)^{n-1}$ different parking functions on the set $[n]$. Since then, a substantial theory of parking functions has been developed, with links to trees and priority queues (Gilbey and Kalikow (1999)), partitions (Stanley (1997)), and representation theory (Haiman (1994)). More recently, generalizations of parking functions have found application in areas of statistical physics like the modelling of percolation (Majumdar and Dean (2002)), the Abelian sandpile model (Postnikov and Shapiro (2004)) and branching processes (Dumitriu, Spencer, and Yan (2003)).

The name *parking function* is motivated by the following process: Suppose that m drivers each choose a preferred parking space in a linear car park with n spaces. Each driver goes to the chosen space and parks there if it is free, and otherwise takes the first available space with a larger number (if any). If all drivers park successfully, the sequence of choices is called a parking function. In general, if k drivers fail to park, we have a *defective parking function* of *defect* k. This concept generalizes further to x*-parking functions* (Pitman and Stanley, 2002), which will not be discussed in this thesis.

In Chapter 3, we establish a recurrence relation for the exact numbers $\mathrm{cp}(n, m, k)$ of parking functions from $[n]$ to $[m]$ of defect k, and express this as an equation for a three-variable generating function. We solve this equation using the kernel method, and extract the coefficients explicitly: it turns out that the cumulative totals are partial sums in Abel's binomial identity.

Finally, we compute the asymptotics of $\mathrm{cp}(n, m, k)$. In particular, for the case $m = n$, we show that if choices are made independently at random, the limiting distribution of the defect (the number of drivers who fail to park), scaled by the square root of n, is the Rayleigh distribution. On the other hand, in the case $m = \omega(n)$, we show that the probability that all spaces are occupied tends asymptotically to one.

Indication of source. The content of Chapter 3 has been previously published in the *Electronic Journal of Combinatorics* (Cameron, Johannsen, Prellberg, and Schweitzer (2008)).

1.2. Randomized Search Heuristics

The second part of this thesis addresses our second question. We perform runtime analyses of a randomized search heuristic on a number of classical problems in combinatorial optimization.

Mathematically, an optimization problem can be modeled by a search space containing all potential solutions and an objective function on this space. To optimize this function, a randomized search heuristic successively generates random candidate solutions according to some distribution. This distribution usually depends on the objective values of the previously generated solutions and some kind of neighborhood structure on the search space.

For example, consider the randomized search heuristic *Randomized Local Search*. For this heuristic, search points are considered to be vertices in a finite and connected graph such that the edges of the graph define the neighborhood structure of the search space. Starting with an arbitrary candidate solution, the candidate solution generated by Randomized Local Search in each subsequent iteration is generated by choosing a random neighbor of the best candidate solution seen so far.

We are interested in the optimization times of a given randomized search heuristic on different optimization problems. The *optimization time* is the random variable that counts the number of candidate solutions to be tested until a solution of optimal objective value is found.

The optimization time of a randomized search heuristic on a given problem can be infinite. For example, Randomized Local Search finds a local optimum with probability one but is not able to find a better search point afterwards. Thus, unless the first local optimum found by Randomized Local Search is also a global optimum, the expected optimization time is infinite.

During the last two decades the analysis of the optimization time of randomized search heuristics has become a growing research field and seminal results have be shown for Randomized Local Search (Papadimitriou, Schäffer, and Yannakakis (1990)), simulated annealing (Sasaki and Hajek (1988)), and evolutionary algorithms (Beyer, Schwefel, and Wegener (2002); Droste, Jansen, and Wegener (2002); He and Yao (2002)).

In the last years, combinatorial optimization problems became the benchmark of theoretical research on evolutionary algorithms (see Chapter 5). On the one hand, these problems are general enough to make meaningful comparisons among different evolutionary algorithms. On the other hand, combinatorial optimization problems have enough structural properties to make the theoretical analysis of such algorithms possible.

Among evolutionary algorithms, the optimization time has been studied most for the (1+1) Evolutionary Algorithm. In the terminology above, the (1+1) Evolutionary Algorithm is a variant of Randomized Local Search. Instead of choosing a random neighbor of the currently best candidate solution, we generate a new candidate solution by performing a random walk. The number of steps in this walk is given by the Poisson distribution with parameter one[1]. Consequently, on average, the random walk performs a single step which corresponds to the behavior of Randomized Local Search. However, in every iteration each point of the search space has a positive probability to be the candidate solution of the (1+1) Evolutionary Algorithm. Therefore, in contrast to Randomized Local Search, the expected optimization time of the (1+1) Evolutionary Algorithm is always bounded by a function of the size of the search space.

The content of the second part is split in two chapters. The first chapter (Chapter 4) introduces probabilistic techniques for the analysis of randomized search heuristics. The second chapter (Chapter 5) applies these techniques to the runtime analysis of evolutionary algorithms.

1.2.1. Probabilistic Methods for Randomized Search Heuristics

In Chapter 4, we introduce the probabilistic techniques as used in the analysis of the evolutionary algorithms in Chapter 5 and put them into the context of related techniques for the analysis of randomized search heuristics. The main focus of Chapter 4 is on drift analysis. We further discuss dominance of stochastic precesses and the random process commonly referred to as the *gambler's ruin*.

Indication of source. The content of Chapter 4 contains parts that have been previously published in the *Proceedings of GECCO '08* (Happ, Johannsen, Klein, and Neumann (2008)) and the *Proceedings of GECCO '10* (Doerr, Johannsen, and Winzen (2010b)) and results that will appear in the *Proceedings of CEC '10* (Doerr, Johannsen, and Winzen (2010a)).

1.2.2. Evolutionary Computation in Combinatorial Optimization

Chapter 5 has been written for the purpose of being a survey included in the book *Theory of Randomized Search Heuristics* (Auger and Doerr (2010)) and is presented as such. This chapter gives a concise overview over the runtime analysis of the (1+1) Evolutionary Algorithm on polynomially solvable combinatorial optimization problems. In the spirit of our first guiding question to study different problems for the same search

[1] On the search space of n-bit vectors we flip each bit independently with probability $1/n$ instead of performing a random walk. However, conceptually this is not significant as the binomial distribution converges towards the Poisson distribution.

heuristic, the main focus of this chapter is to present the discussed problem specifications, algorithms, and solution methods in a unified and consistent way.

Indication of source. The content of Chapter 5 will appear in the book *Theory of Randomized Search Heuristics* (Auger and Doerr (2010)). Chapter 5 contains results that have been previously published in the *Proceedings of GECCO '07* (Doerr and Johannsen (2007b)) and in the *Proceedings of GECCO '10* (Doerr and Johannsen (2010)).

1.3. Further Contributions

Not all of the author's work fits the structure of this thesis. In Appendix A, we list further contributions that are part of the authors PhD research but that are not further discussed in this thesis.

Part I

Random Combinatorial Structures

2
Vertex Degrees in Random Planar Maps

A *map* is a combinatorial embedding of a connected planar graph to the sphere, where generally multiple edges and loops are admissible[1]. We can completely characterize a map by its underlying (multi-)graph, together with a cyclic ordering of the edges around each vertex or, equivalently, by the sets of it vertices, edges, and faces. Following standard definitions, we say that a map is *biconnected*, if its edge set cannot be partitioned into two non-empty subsets, such that there is only one vertex incident with edges from both sets. We say that a map is *3-connected*, if the underlying planar graph is 3-connected and has neither loops nor multiple edges. We shall refer to rooted 3-connected maps as *c-nets*, since this is their common name in the literature.

The study of maps has a long history. Already Euler asked for the number of isomorphism types of convex polyhedra (Federico (1975)), which, by a well-known theorem of Steinitz (1922), are combinatorially equivalent to 3-connected planar graphs. Subsequently, Whitney (1932) showed that all combinatorial embeddings of such a graph are topologically equivalent, thus implying the existence of a simple one-to-one correspondence between c-nets and 3-connected planar graphs. However, Euler's question still remains unanswered.

A general theory of map enumeration was initiated by Tutte (1963) in the early 60's, who studied systematically the number of maps in several fundamental classes. Since then, maps have been investigated extensively as combinatorial as well as geometric objects, and a rich theory highlighting several properties and aspects of maps has evolved. We mention here selectively two recent results. First, Fusy, Poulalhon, and Schaeffer (2005) discovered a beautiful bijection between the class of c-nets and a class of plane trees, which not only provides a combinatorial interpretation of the formula enumerating c-nets with a given number of vertices and faces, but it also solves the problem of compressing efficiently the connectivity information encoded in such a map. Second, Aleardi, Devillers, and Schaeffer (2006) gave, among other results, optimal

[1]We use the two terms *planar map* and *map* synonymously.

representations for several classes of maps and triangulations.

We now advance to the question about "typical" properties of maps. This means that we ask for structural properties of random maps (for example, maps drawn uniformly at random from, say, the set of all maps with n edges) that are observed with high probability, that is, with probability tending to one as $n \to \infty$. Not much is known about random maps. One reason for this lack of understanding is that maps are heavily *constrained* combinatorial objects, in the sense that the appearance of specific edges highly depends on the presence or absence of other edges. Therefore, we resort to exact counting techniques to obtain precise results. One aim of this chapter is to attack precisely this problem and to demonstrate that maps contain in a well-defined sense enough "independence", allowing us to study their typical asymptotic properties by using well-established methods from classical random graph theory.

Following the standard approach in the literature, we consider *rooted* maps, that is, maps with a distinguished oriented edge called the *root*. Note that the direction of the root edge implies the existence of a distinguished *root vertex*. Moreover, together with the orientation of the sphere, the direction of the root edge also implies the existence of a distinguished *root face* incident with the root edge. This common restriction has the advantage of greatly simplifying the analysis, without affecting statistical properties, since a result by Richmond and Wormald (1995) implies that almost all large maps are asymmetric. From now on, all considered maps are rooted objects.

For $n \in \mathbb{N}$, let $\mathcal{F}_n$ be a class of maps with n edges, and denote by F_n a map that is drawn uniformly at random from this set. Our main objective in this chapter is to study the number $\deg(k;\, \mathsf{F}_n)$ of vertices of degree k in F_n. In the special case that $\mathcal{F}_n$ is the class $\mathcal{G}_n$ of all (*general*) maps with n edges, Gao and Wormald (2003) determined an asymptotic expression for the expected value $\mathrm{E}[\deg(k;\, \mathsf{F}_n)]$, and showed accompanying concentration statements that become sharp as the size n of the maps tends to infinity.

Before we state this result we introduce some additional notation. For a function $F(x)$ that has a Taylor expansion of $F(x)$ around $x = 0$ we denote by $[x^k]F(x)$ the coefficient of x^k in this Taylor expansion. By $(1 \pm \varepsilon)\, N$, we denote the open interval $((1-\varepsilon)N, (1+\varepsilon)N)$. We denote by $\alpha(n) \sim_n \beta(n)$ that $\lim_{n\to\infty} \alpha(n)/\beta(n) = 1$.

Theorem 2.1 (Gao and Wormald (2003)). *Let $\varepsilon > 0$, $k \in \mathbb{N}$. Let G_n be a map drawn uniformly at random from the class $\mathcal{G}_n$ of all maps with n edges. Then, for sufficiently large n*

$$\Pr[\deg(k;\, \mathsf{G}_n) \in (1 \pm \varepsilon)\, g_k\, n] \geq 1 - \mathcal{O}(\varepsilon^{-2} (g_k\, n)^{-1} \log^{20} n)\,, \tag{2.0.1}$$

where $g_k = [z^k] D_{\mathcal{G}}(z)$ and $D_{\mathcal{G}}(z) = \frac{1}{4}\left(\sqrt{\frac{6+3z}{6-5z}} - 1\right)$. Furthermore,

$$g_k \sim_k \frac{1}{\sqrt{10\pi}}\, k^{-1/2} \left(\tfrac{5}{6}\right)^k .$$

In Gao and Wormald (2003) the quantities g_k were given in a slightly different form. However, simple algebraic manipulations lead to the above expression, which is more suitable for our intended application.

If we proceed to more complex classes of maps far less is known. In this context, Liskovets (1999) determined for the classes of biconnected, Eulerian and loopless maps

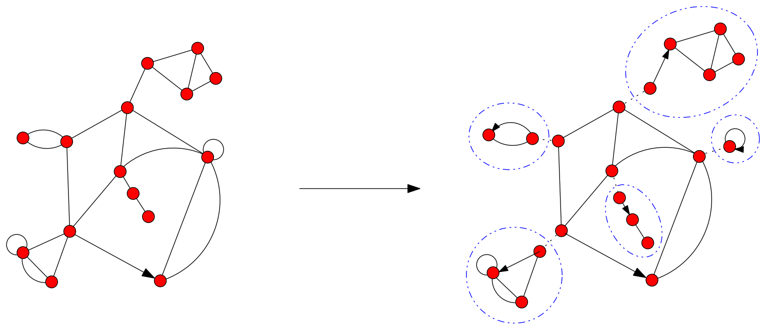

Figure 2.1. Decomposition of a general map into a biconnected map and attached submaps. The roots are indicated by arrows.

as well as for bi- and 3-connected triangulations the limiting probability that a vertex has degree k. This result can be used to derive the expected number of vertices of degree k in a random map from the corresponding class; however, it does not provide any other information about the underlying distribution.

Our Results Our main contribution is a universal framework, which allows us to directly transfer concentration results concerning the total number of vertices of degree k in a random map from a class $\mathcal{M}$ to concentration results concerning a random map from another class $\mathcal{C}$, which depends in a suitable way on $\mathcal{M}$. In particular, we consider so-called *composition schemata*, where "simple" classes of maps are constructed out of maps that have a "higher" complexity (for example, higher connectivity). Let us mention one example. Any general map decomposes uniquely into the maximal biconnected submap containing the root, and a set of other maps, which are each attached to B at a single vertex, see Figure 2.1. So, any general map can be (recursively) constructed from a biconnected map by replacing each vertex of the biconnected map with an appropriate set of attachment maps. Using our framework, we exploit this relationship and derive for random biconnected maps the following theorem as a consequence of and a counterpart to Theorem 2.1.

Theorem 2.2. *Let $\varepsilon > 0$, $k \in \mathbb{N}$. Let B_n be a map drawn uniformly at random from the class $\mathcal{B}_n$ of all biconnected maps with n edges. Then, for sufficiently large n*

$$\Pr[\deg(k;\, \mathsf{B}_n) \in (1 \pm \varepsilon)\, b_k\, n] = 1 - o(1)\,, \tag{2.0.2}$$

where

$$b_k \sim_k \tfrac{3}{\sqrt{2\pi}}\, k^{-1/2} \left(\tfrac{2}{3}\right)^k .$$

It turns out that many other important classes of maps, like c-nets, loop- and bridgeless maps, and several classes of triangulations, can be described by appropriate composition schemata, see the very detailed survey in Banderier et al. (2001). Our main result (Theorem 2.12) addresses precisely those classes, and asserts essentially that for all of them we can derive appropriate concentration results for the number of vertices of degree k. In Section 2.4, we prove the previous theorem. Equivalent results for random loopless and bridgeless maps are given in Section 2.6 . We refer the reader to Section 2.1 and Section 2.2 for a formal description of the properties of the maps that we consider, and to Section 2.3 for our general framework.

We conclude the chapter with a concentration result on the degree sequence of random c-nets (that is, 3-connected maps). Recall that by Whitney's theorem the classes of c-nets and 3-connected planar graphs coincide. Thus, the following theorem, which is proven in Section 2.5 is also the first non-trivial result about the distribution of the number of vertices of degree k in random 3-connected planar graphs.

Theorem 2.3. *Let $\varepsilon > 0$, $k \in \mathbb{N}$. Let T_n be a map drawn uniformly at random from the class $\mathcal{T}_n$ of all c-nets with n edges. Then, uniformly for sufficiently large n*

$$\Pr[\deg(k;\, \mathsf{T}_n) \in (1 \pm \varepsilon)\, t_k\, n] = 1 - o(1)\,, \tag{2.0.3}$$

where

$$t_k \sim_k \tfrac{9\sqrt{3}}{\sqrt{2\pi}}\, k^{-1/2}\, 2^{-k}\,.$$

Techniques Let us discuss on a high level our proof strategy. We shall restrict this exposition to the special case of Theorem 2.2; the main line of reasoning for Theorem 2.12 is very similar. As a starting point for our argument we use the results in Banderier et al. (2001), where, the authors showed that with probability $\Theta(n^{-2/3})$ a random map G on $3n$ edges contains one giant maximal biconnected submap B on exactly n edges (B is called the *core* of G). In other words, G decomposes into B and $2n$ possibly empty (that is, consisting of no edges) maps $G_1, G_2, \ldots, G_{2n}$, each of which is combinatorially embedded in a face of B, attached to B at one vertex, and disjoint from the remaining G_i's otherwise (recall Figure 2.1).

Since the total number of edges contained in the G_i's is $|G|-|B| = 2n$, each G_i has on average exactly one edge. Intuitively, this suggests that the single G_i's shouldn't contain too many edges. We can confirm this intuition and show that with exponentially high probability all but $o(n)$ many of the G_i's are "small". Since any pair of maps G_i and G_j intersect at most in the single vertex they share with B, we then infer by using large deviation bounds (see Chapter 4) that the total number of vertices of degree k that are in the G_i's but not in B is extremely sharply concentrated. What remains is to handle the vertices in B.

In the next step of our proof we show an important stochastic property of G. More specific, we argue that we can "generate" a random map G by first choosing uniformly at random a core B of size n, and then choosing independently the maps G_i according to the *Boltzmann distribution*. Unfortunately, the Boltzmann distribution cannot guarantee that the resulting map G is of size $3n$. However, we can condition on the event "$|G| = 3n$" at a multiplicative loss of $\Theta(n^{2/3})$ to the probability bounds we want to obtain. Compared to the order of bounds we derive afterwards, this loss is acceptable.

Having this fact, completing the proof is routine: as in the Boltzmann model everything behaves independently, the relevant random variables are extremely sharply concentrated. Then, by assuming that the number of vertices of degree k is not concentrated for random biconnected maps, we infer that a similar non-concentration result must be true for general maps — a contradiction to Theorem 2.1.

Outline This chapter is structured as follows. Section 2.1 deals with basic facts from the enumerative theory of maps and introduces the relevant known results for random maps. In Section 2.2 we describe the Boltzmann model, tailored to our specific application. Subsequently, in Section 2.3, we introduce our framework for transferring concentration results between different map classes and present in Theorem 2.12 our main result. Finally, in Sections 2.4, 2.5, and 2.6 we apply the framework to random biconnected, 3-connected, loopless and bridgeless maps.

Basic Notation For any map M we write $V(M)$ for the set of its vertices and $E(M)$ for the set of its edges. Each edge can be *marked* or not. In our case, all edges in $E(M)$ are marked with the possible exception of the root, which may or may not be marked. We denote maps where the root is not marked by M°, that is, the only difference between M and M° is that in M the root is marked and in M° it is not. The *size* $|M|$ of M is equal to the number of marked elements in $E(M)$; thus, $|M^\circ| = |M| - 1$. Finally, the degree of the root vertex of M is denoted by $\mathrm{rdeg}(M)$. Again, only marked edges contribute to the degree of a vertex, that is, $\mathrm{rdeg}(M^\circ) = \mathrm{rdeg}(M) - 1$.

Let $\mathcal{F}$ be any class of maps for which the roots are marked. By $\mathcal{F}^\circ$ we denote the maps obtained from the maps in $\mathcal{F}$ by unmarking the root, with the exception of the two maps on one edge (the single loop "$\bullet\!\circlearrowright$" and the single edge "$\bullet\!\rightarrow\!\bullet$"), which are never elements of $\mathcal{F}^\circ$. By $\mathcal{F}_n$ and $\mathcal{F}_n^\circ$ we denote the subsets of maps in $\mathcal{F}$ and $\mathcal{F}^\circ$ that have exactly n marked edges (including and excluding the root, respectively).

For a class $\mathcal{F}$ of maps, we write $F(x)$ and $F(x,z)$ for the ordinary generating functions enumerating all maps in $\mathcal{F}$, where x marks the number of edges and z the root-degree, that is, $F(x) = \sum_{M\in\mathcal{F}} x^{|M|}$ and $F(x,z) = \sum_{M\in\mathcal{F}} x^{|M|} z^{\mathrm{rdeg}(M)}$. Let $z > 0$. By $\rho_{\mathcal{F}}$ we denote the dominant singularity of $F(x,z)$ with respect to x. Note that in general $\rho_{\mathcal{F}}$ may depend on z; however, in all classes of maps considered here this is not the case.

2.1. Map Compositions and the Cores of Random Maps

In a seminal work, Tutte (1963) laid the basis of an enumerative theory of maps, where he determined exact formulas for the number of general maps, biconnected maps and c-nets on a given number of edges. Since then this theory has been largely developed and extended, revealing deep insights in the combinatorial structure and properties of maps. In this section we introduce some basic facts about the enumeration of maps. Moreover, we present the notions and concepts given by Banderier et al. (2001) which we use to develop the methods in Section 2.3 and to obtain the results in Sections 2.4.

Let $\mathcal{M}$ be a class of maps and $\mathcal{C}$ a subset of $\mathcal{M}$ defined by additional properties (typically, higher connectivity or more complex structure). We say that the class $\mathcal{C}$ is the class of *core* maps of $\mathcal{M}$ with respect to the class of *substitution* maps $\mathcal{H}$, if for all elements $M \in \mathcal{M}$ there is at most one element $C \in \mathcal{C}$ such that M can be composed by substituting in a unique way all edges of C by maps in $\mathcal{H}$ as follows[2]. First, we assign directions to all edges of C (this is done canonically with respect to the direction of the root edge of C). Second, we substitute all marked edges of C by maps in $\mathcal{H}$ such that the roots of those replace the edges of C while respecting their direction. Finally, we replace all former root edges of the substituted maps by ordinary undirected and marked edges if they were marked before or remove them otherwise (if the root was substituted, the root of the substitution map becomes the new root). If for an $M \in \mathcal{M}$ there is a $C \in \mathcal{C}$ with the above properties, then we write $C(M) = C$, and $C(M) = \perp$ otherwise.

Following the symbolic method as in Flajolet and Sedgewick (2009), we can describe this composition by the schema

$$\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}, \tag{2.1.1}$$

where $\mathcal{D}$ is the subclass of maps in $\mathcal{M}$ that have an empty core, that is, for any $M \in \mathcal{D}$ we have $C(M) = \perp$. Here, "$\circ$" represents precisely the edge substitution described above, and "$+$" denotes the disjoint union of two combinatorial classes. Let M be a map in $\mathcal{M}$ with a core $C = C(M)$ and let the corresponding set of substitution maps be $\mathcal{H}(M) = \{H_1, \ldots, H_{|C|}\}$. We then write $M = C \circ (H_1, \ldots, H_{|C|})$ with slight abuse of notation.

The schema (2.1.1) directly translates to a relation for the corresponding generating functions,

$$M(x) = C(H(x)) + D(x), \tag{2.1.2}$$

where x marks the edges of the maps. The generating functions of the map classes we study show common analytic properties.

Definition 2.4. *Let F be a generating function that is analytic at $x = 0$ and has radius of convergence $\rho_{\mathcal{F}}$. Then F is called* ***singular with exponent 3/2*** *if there exist positive constants ε, f_0, f_1, $f_{3/2}$ such that*

(i) $F(x)$ *is analytic on all* $x \in \mathbb{C}$ *for which* $|x| = \rho_{\mathcal{F}}$ *and* $x \neq \rho_{\mathcal{F}}$*;*

(ii) $F(x)$ *is continuable in* $\Delta = \{x \in \mathbb{C}\colon |x| < \rho_{\mathcal{F}} + \varepsilon \text{ with } x \notin [\rho_{\mathcal{F}}, \rho_{\mathcal{F}} + \varepsilon]\}$*;*

(iii) $F(x) = f_0 + f_1(1 - x/\rho_{\mathcal{F}}) + f_{3/2}(1 - x/\rho_{\mathcal{F}})^{3/2} + \mathcal{O}\big((1 - x/\rho_{\mathcal{F}})^2\big)$ *as* $x \to \rho_{\mathcal{F}}$ *in* Δ.

For example, the generating functions for the classes of general maps, biconnected maps and c-nets are singular with exponent 3/2. Table 2.1 summarizes the respective constants for those classes and for some other classes that become relevant in Section 2.4. If the generating function of a class of maps $\mathcal{F}$ is singular with exponent 3/2, then the following statement allows us to determine an asymptotic expression for its coefficients.

Theorem 2.5 (see e. g., Corollary VI.1 of Flajolet and Sedgewick (2009)). *Assume that $f(z)$ is analytic in $\Delta := \Delta(\phi, R) = \{z : z \neq 1, |z| < R, |\arg(z - 1)| > \phi\}$, where*

[2] For the example in the introduction, we replace each edge by a new edge with two pending maps, thus effectively attaching maps to the vertices of the core.

$\mathcal{F}$	class of maps	ρ_F	f_0	f_1	$f_{3/2}$
$\mathcal{G}$	general maps	$\frac{1}{12}$	$\frac{1}{3}$	$-\frac{4}{3}$	$\frac{8}{3}$
$\mathcal{B}$	non-separable (biconnected) maps	$\frac{4}{27}$	$\frac{1}{3}$	$-\frac{4}{9}$	$\frac{8\sqrt{3}}{81}$
$\mathcal{T}^\circ$	3-connected maps (c-nets), root unmarked	$\frac{1}{4}$	$\frac{1}{540}$	$-\frac{167}{8100}$	$\frac{32}{729}$
$\mathcal{L}$	loopless / bridgeless	$\frac{27}{256}$	$\frac{32}{27}$	$-\frac{32}{81}$	$\frac{32\sqrt{6}}{81}$

Table 2.1. Parameters for the singular expansions corresponding to different map classes (see also Banderier et al. (2001)).

$R > 1$ and $0 < \phi < \pi/2$. If, as $z \to 1$ in Δ, $f(z) = (1-z)^{-\alpha} + o((1-z)^{-\alpha})$ holds for some $\alpha \notin \{0, -1, -2, \dots\}$, then

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)},$$

where $\Gamma(x)$ denotes the Gamma function.

For example, we obtain precise asymptotic estimates for the quantities $|\mathcal{F}_n|$.

Theorem 2.6. *Let $\mathcal{F}$ be a class of maps for which the corresponding generating function $F(x)$ is singular with exponent $3/2$. Then*

$$|\mathcal{F}_n| \sim \frac{3f_{3/2}}{4\sqrt{\pi}}\, n^{-5/2}\, \rho_{\mathcal{F}}^{-n}.$$

With all the above facts at hand we are ready to define the properties of the composition schemata that are of interest in the remainder of this chapter.

Definition 2.7. *Let $\mathcal{M}$, $\mathcal{C}$, $\mathcal{H}$, and $\mathcal{D}$ be classes of maps related by the composition schema $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$. We say this schema is of singular type $(3/2 \circ 3/2)$ if the generating functions C and H are singular with exponent $3/2$. We say it is critical if $\rho_M = \rho_C = H(\rho_H)$. Finally, we call it a* ***proper map composition schema*** *if it is of singular type $(3/2 \circ 3/2)$, critical, and $\mathcal{H}$ is closed under inversion of the orientation of the root edge. In addition, neither the root of any map in $\mathcal{H}$ nor the marked edges of any map in $\mathcal{C}$ are allowed to be loops.*

Random maps from classes that are related through proper map decomposition schemata have been studied extensively by Gao and Wormald (1999) and by Banderier et al. (2001). In particular, in Banderier et al. (2001) a precise characterization of the probability that a random map has a core of a given size. Here we state a suitable special case of this far more general result, tailored to our specific application.

Theorem 2.8 (Theorem 5 from Banderier et al. (2001)). *Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ be a proper map composition schema. Moreover, set $c_{\mathcal{H}} = -h_1/h_0$, where h_0 and h_1 are the first two coefficients in the singular expansion of $H(x)$, and let $\mathsf{M}_{\lceil c_{\mathcal{H}} n \rceil}$ be drawn uniformly at random from $\mathcal{M}_{\lceil c_{\mathcal{H}} n \rceil}$. Then there exists a constant $c > 0$ such that for* *large* n

$$\Pr\left[C(\mathsf{M}_{\lceil c_{\mathcal{H}} n\rceil}) \in \mathcal{C}_n\right] \sim c\, n^{-2/3}\,.$$

2.2. Random Maps in the Boltzmann Model

All classes of maps considered in this chapter allow a so-called decomposition, which is a (recursive) description in terms of other classes of higher complexity. One substantial benefit of such a decomposition is that it enables us to mechanically develop algorithms that sample maps from the class in question by using the framework of *Boltzmann samplers*. Such sampling algorithms are an important ingredient in our proofs, and were used for the first time systematically by Bernasconi et al. (2008) to study properties of random structures.

The Boltzmann model was introduced by Duchon et al. (2004). Let $\mathcal{F}$ be any class of maps and $F(x)$ the corresponding generating function. The *Boltzmann distribution* $\Gamma F(x)$ with parameter x assigns to each map $M \in \mathcal{F}$ the probability

$$\Pr[M] = \frac{x^{|M|}}{F(x)}, \tag{2.2.1}$$

if this expression is well-defined. Note that the above probability depends just on $|M|$ if x is fixed; hence, all maps of the same size have the same probability of being drawn. In other words, $\Gamma F(x)$ is uniform for each size n. In the following section we see that this model turns out to be useful in proving properties of maps drawn uniformly at random from $\mathcal{F}_n$.

Let F be a random map from $\mathcal{F}$ drawn according to the Boltzmann distribution $\Gamma F(\rho_{\mathcal{F}})$, that is, with the dominant singularity $\rho_{\mathcal{F}}$ of $F(x)$ as parameter. Here we silently assume that $F(\rho_{\mathcal{F}})$ is finite, which is the case for all classes considered in this thesis, see Table 2.1. We denote by $P_{\mathcal{F}}(x)$ the probability generating function for the size of F, by $R_{\mathcal{F}}(z)$ the probability generating function for the degree of the root vertex of F, and by $E_{\mathcal{F}}(z)$ the function whose k-th coefficient is the expected number of vertices of degree k in F. These functions can be expressed in terms of the univariate and bivariate generating functions $F(x)$ and $F(x,z)$. Recall that $[x^n]F(x)$ is the number of maps in $\mathcal{F}$ with n edges and $[x^n z^k]F(x,z)$ is the number of maps in $\mathcal{F}$ with n edges and root degree equal to k. In particular, $F(x) = F(x,1)$.

Proposition 2.9. *Let $\mathcal{F}$ be a class of maps and F be a random map drawn from $\mathcal{F}$ according to the Boltzmann distribution $\Gamma F(\rho_{\mathcal{F}})$. Then,*

$$P_{\mathcal{F}}(x) := \sum_{n\ge 0} \Pr\left[|\mathsf{F}| = n\right] x^n = \frac{F(\rho_{\mathcal{F}} x)}{F(\rho_{\mathcal{F}})}$$

$$R_{\mathcal{F}}(z) := \sum_{k\ge 0} \Pr\left[\mathrm{rdeg}(\mathsf{F}) = k\right] z^k = \frac{F(\rho_{\mathcal{F}}, z)}{F(\rho_{\mathcal{F}})}\,.$$

Moreover, if $\mathcal{F}$ is closed under re-rooting (where re-rooting means that we replace the root edge of a map by an ordinary edge and then choose one of the other edges to be *the new root edge with an arbitrary direction), then*

$$E_{\mathcal{F}}(z) := \sum_{k\geq 0} \mathrm{E}\big[\deg(k;\mathsf{F})\big]\, z^k = \frac{2\,\rho_{\mathcal{F}}}{F(\rho_{\mathcal{F}})} \int_0^z \frac{1}{t} \cdot \frac{\partial(F(x,t)-F(x,0))}{\partial x}\bigg|_{x=\rho_{\mathcal{F}}} dt + \frac{F(0)}{F(\rho_{\mathcal{F}})}\,.$$

Proof. The first two equations follow directly from the definitions of $F(x)$, $F(x,z)$, and that of the Boltzmann distribution $\Gamma F(\rho_{\mathcal{H}})$. For the third equation, let $n,k \in \mathbb{N}$ and let F_n be chosen uniformly at random from $\mathcal{F}_n$. By double counting in F_n all pairs (v,e), where v is a vertex of degree k, and e is an edge incident with v, we obtain the relation

$$2n \cdot \Pr[\mathrm{rdeg}(\mathsf{F}_n) = k] = k \cdot \mathrm{E}[\deg(k;\mathsf{F}_n)]$$

and again apply the definition of $F(x)$ and $F(x,z)$. The last term is obtained by considering the special case $k=0$. □

For a map class $\mathcal{F}$ whose associated generating function is singular with exponent $3/2$, the next statement gives asymptotic estimates for probabilities of certain events in the Boltzmann model.

Corollary 2.10. *Let $n \in \mathbb{N}$. Let $\mathcal{F}$ be a class of maps that is singular with exponent $3/2$ and let F be a random map from $\mathcal{F}$ drawn according to the Boltzmann distribution $\Gamma F(\rho_{\mathcal{F}})$. Set $d_{\mathcal{F}} = \frac{3f_{3/2}}{4\sqrt{\pi}\,f_0}$. Then, $\mathrm{E}\big[|\mathsf{F}|\big] = -f_1/f_0$ and furthermore*

$$\begin{aligned} \Pr\big[|\mathsf{F}| = n\big] &\sim d_{\mathcal{F}}\, n^{-5/2}\,, \\ \Pr\big[|\mathsf{F}| \geq n\big] &\sim \tfrac{2}{3}\, d_{\mathcal{F}}\, n^{-3/2}\,, \\ \sum_{k\geq n} k \cdot \Pr\big[|\mathsf{F}| = k\big] &\sim 2\, d_{\mathcal{F}}\, n^{-1/2}\,. \end{aligned}$$

Proof. The first statement follows directly from Theorem 2.6. The second statement can then be derived from the first statement, since

$$\Pr\big[|\mathsf{F}| \geq n\big] = \sum_{k\geq n} \Pr\big[|\mathsf{F}| = n\big] \sim \sum_{k\geq n} d_{\mathcal{F}}\, k^{-5/2} \sim d_{\mathcal{F}} \int_n^\infty t^{-5/2} dt = \tfrac{2}{3}\, d_{\mathcal{F}}\, n^{-3/2}\,.$$

Similarly, the third statement follows, since

$$\sum_{k\geq n} k \cdot \Pr\big[|\mathsf{F}| = k\big] \sim \sum_{k\geq n} d_{\mathcal{F}}\, k^{-3/2} \sim d_{\mathcal{F}} \int_n^\infty t^{-3/2} dt = 2\, d_{\mathcal{F}}\, n^{-1/2}\,.$$

□

The main strength of the Boltzmann model is that composition schemata of combinatorial classes can be translated into a relation of the corresponding Boltzmann distributions. This gives rise to efficient algorithms to generate objects according to the distribution $\Gamma F(x)$, called *Boltzmann samplers*. Duchon et al. (2004), and moreover Fusy (2005), gave several general procedures which translate common combinatorial construction rules like union, set, substitution etc. into Boltzmann samplers. Here, we only need the relation of between the Boltzmann distributions ΓM, ΓC, and ΓH given by the composition schema $\mathcal{M} = \mathcal{C} \circ \mathcal{H}$.

Lemma 2.11 (Fusy (2005)). *Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H}$ be a composition schema for a class of maps. Let M be a map in $\mathcal{M}$ such that $M = C \circ (H_1, \dots, H_{|C|})$ where the core C is drawn from $\mathcal{C}$ according to the Boltzmann distribution $\Gamma C(H(x))$ and the substitution maps $(H_i)_{1 \le i \le |C|}$ are drawn independently from $\mathcal{H}$ according to the Boltzmann distribution $\Gamma H(x)$. Then the distribution of M is the Boltzmann distribution $\Gamma M(x)$.*

In the next section we see how to make use of the previous statement. On the one hand, in the Boltzmann model, a random map from $\mathcal{M}$ is composed of a number of maps chosen independently from $\mathcal{H}$. On the other hand, if we condition on the event that a map in the Boltzmann model has a specific size, then it is uniformly distributed over all maps of that size. This combination allows us in the next section to condition on suitable events in the uniform model while still exploiting the independence of the Boltzmann model.

2.3. Degree Inheritance for Large Cores

Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ be a proper map composition scheme. Let $k \in \mathbb{N}$ be fixed, let $n \in \mathbb{N}$ be sufficiently large, and let $m = \lceil c_{\mathcal{H}}\, n \rceil$ with $c_{\mathcal{H}}$ as in Theorem 2.8. In this section we show that if the number of vertices of degree k in a map M_m drawn uniformly at random from $\mathcal{M}_m$ is concentrated around its expectation, then so is the number $\deg(k;\, \mathsf{C}_n)$, where C_n is chosen uniformly at random from $\mathcal{C}_n$. In particular, we show that if there exists a function $D_{\mathcal{M}}(z)$ such that

$$\deg\big(k;\, \mathsf{M}_m\big) \in (1 \pm \varepsilon)\, [z^k] D_{\mathcal{M}}(z)\, m \tag{2.3.1}$$

with probability $1 - o\left(n^{-2/3}\right)$, then there exists a function $D_{\mathcal{C}}(z)$ such that

$$\deg\big(k;\, \mathsf{C}_n\big) \in (1 \pm \varepsilon)\, [z^k] D_{\mathcal{C}}(z)\, n \tag{2.3.2}$$

with probability $1 - o(1)$.

In order to show (2.3.2) first recall that any map $M \in \mathcal{M}$ that has a non-empty core can be represented as $M = C \circ (H_1, \dots, H_{|C|})$, where the core of M is C, and the substitution maps $(H_i)_{1 \le i \le |C|}$ replace the marked edges of C. Our main strategy splits the vertices of degree k in M into two sets and counts them separately. The first set contains all vertices of degree k which lie in one of the maps $(H_i)_{1 \le i \le |C|}$ but not in C (that is, this set contains all non-root vertices of degree k in the H_i's). More formally, set

$$\operatorname{adeg}(k;\, M) := \left|\{v \in V(M) \setminus V(C(M))\colon \deg(v; M) = k\}\right|.$$

The second set contains all vertices of degree k that lie in the core C of M. These vertices can have neighbors inside the core and outside the core. To obtain the desired concentration results, we have to distinguish them further by their degree ℓ in C. We set

$$\operatorname{bdeg}(k, \ell;\, M) := \big|\{v \in V(C)\colon \deg(v;\, M) = k \text{ and } \deg(v;\, C) = \ell\}\big|.$$

Using these definitions we infer that

$$\deg(k;\, M) = \operatorname{adeg}(k;\, M) + \sum_{\ell=0}^{k} \operatorname{bdeg}(k,\ell;\, M). \tag{2.3.3}$$

Now suppose that (2.3.1) holds and that the core $C(\mathsf{M}_m)$ of M_m is in $\mathcal{C}_n$, that is, contains precisely n edges. Recall, that by Theorem 2.8 this happens with probability $\Theta(n^{-2/3})$. Note that in this case $C(\mathsf{M}_m)$ is uniformly distributed in $\mathcal{C}_n$, and so we can write $\mathsf{M}_m = \mathsf{C}_n \circ (H_1, \dots, H_n)$, where C_n is a map drawn uniformly at random from the class $\mathcal{C}_n$, and the $(H_i)_{1\le i\le |C|}$ are random maps from $\mathcal{H}$ such that $\sum_{i=1} |H_i| = \lceil c_{\mathcal{H}}\, n\rceil$.

Let us look a little closer at the quantities $\operatorname{adeg}(k;\, \mathsf{M}_m)$ and $\operatorname{bdeg}(k,\ell;\, \mathsf{M}_m)$. First, $\operatorname{adeg}(k;\, \mathsf{M}_m)$ enumerates all non-root vertices in the H_i's that have degree k. Moreover, the H_i's contain in total $\lceil c_{\mathcal{H}}\, n\rceil$ edges, which means that the average number of edges (and consequently also the average number of vertices) in each H_i is in $\mathcal{O}(1)$. So the impact of a typical H_i on $\operatorname{adeg}(k;\, \mathsf{M}_m)$ is not too large. This allows us to prove that there is a quantity a_k such that with high probability

$$\operatorname{adeg}(k;\, \mathsf{M}_m) \in (1 \pm \varepsilon)\, a_k\, n. \tag{2.3.4}$$

Let us set $A(z) = \sum_{i\ge 0} a_i z^i$. In a similar way we can think about $\operatorname{bdeg}(k,\ell;\, \mathsf{M}_m)$, which counts the number of vertices in $\mathsf{C}_n = C(\mathsf{M}_m)$ that have degree k in M_m, and degree $\ell \le k$ in C_n. Let $v \in \mathsf{C}_n$, and note that $\deg(v;\, \mathsf{M}_m)$ equals the sum of the root degrees of $\deg(v;\, \mathsf{C}_n)$ of the H_i's, namely those that replace the edges of C_n that are incident with v. Again, as the H_i's are in average small, we show that there are quantities $b_{k,\ell}$ such that with high probability

$$\operatorname{bdeg}(k,\ell;\, \mathsf{M}_m) \in (1 \pm \varepsilon)\, b_{k,\ell} \deg(\ell;\, \mathsf{C}_n). \tag{2.3.5}$$

In fact, we show later that there is a function $B(z)$ such that $b_{k,\ell} = [z^k]B(z)^\ell$.

With these considerations at hand, we obtain a recursive definition for the expected values of the quantities $\deg(\ell; \mathsf{C}_n)$ by combining (2.3.1) with (2.3.3)–(2.3.5). More precisely, if we let $D_{\mathcal{C}}(z)$ be the generating function

$$D_{\mathcal{C}}(z) = \frac{1}{n} \sum_{\ell\ge 0} \mathrm{E}[\deg(\ell; \mathsf{C}_n)] z^\ell\,,$$

then the above discussion suggests that

$$c_{\mathcal{H}} [z^k] D_{\mathcal{M}}(z) = [z^k]A(z) + \sum_{\ell=0}^{k} [z^k]B(z)^\ell \cdot [z^\ell] D_{\mathcal{C}}(z)\,,$$

that is,

$$c_{\mathcal{H}} D_{\mathcal{M}}(z) = A(z) + D_{\mathcal{C}}(B(z)).$$

We now show that the above statement is indeed true, thus confirming the obtained intuition. Recall that we denote by $R_{\mathcal{H}}(z)$ the probability generating function for the degree of the root vertex of H, and by $E_{\mathcal{H}}(z)$ the function whose k-th coefficient is the expected number of vertices of degree k in H, where H is a map drawn according to the Boltzmann distribution $\Gamma H(\rho_{\mathcal{H}})$ (see (2.2.1) and Proposition 2.9 in the previous section).

Theorem 2.12. *Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ be a proper map composition schema and let $c_{\mathcal{H}}$ be as in Theorem 2.8. Let $k \in \mathbb{N}$, and let M_n be a map drawn uniformly at random from $\mathcal{M}_n$. Suppose that there exist a generating function $D_{\mathcal{M}}(z)$ and a function $g\colon (0,1) \times \mathbb{N} \to [0,1]$, that is monotone decreasing in both arguments, such that for $\varepsilon > 0$ and large n*

$$\forall\, 0 \le \ell \le k\colon \quad \Pr\left[\deg(\ell; \mathsf{M}_n) \in (1 \pm \varepsilon)\,[z^\ell] D_{\mathcal{M}}(z)\, n\right] \ge 1 - g(\varepsilon, n). \tag{2.3.6}$$

Furthermore, define the function $D_{\mathcal{C}}(z)$ implicitly by

$$c_{\mathcal{H}}\, D_{\mathcal{M}}(z) = \big(E_{\mathcal{H}}(z) - 2\, R_{\mathcal{H}}(z)\big) + D_{\mathcal{C}}\big(R_{\mathcal{H}}(z)\big)\,. \tag{2.3.7}$$

Finally, let the function $h_\alpha\colon (0,1) \times \mathbb{N} \to [0,1]$ with $\alpha(n) \in \omega(1)$ be defined by

$$h_\alpha(\varepsilon, n) = \max\left\{e^{-\varepsilon^2 n/\alpha(n)}, \alpha(n)\, n^{2/3}\, g(\varepsilon/6, n)\right\}.$$

If C_n is drawn uniformly at random from $\mathcal{C}_n$ then for any $\alpha(n) \in \omega(1)$ and $\varepsilon > 0$

$$\Pr\left[\deg(k; \mathsf{C}_n) \in (1 \pm \varepsilon)\,[z^k] D_{\mathcal{C}}(z)\, n\right] \ge 1 - h_\alpha(\varepsilon, n).$$

The remainder of the section is devoted to the proof of the previous theorem. In this, we follow the argument sketched above. In particular, we first show that under the condition $C(\mathsf{M}_m) \in \mathcal{C}_n$, the relations (2.3.4) and (2.3.5) hold with sufficiently high probability.

Proposition 2.13. *Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ be a proper map composition schema. Let $\varepsilon > 0$, $k \in \mathbb{N}$, $\alpha(n) \in \omega(1)$, and n sufficiently large. Furthermore, let $C \in \mathcal{C}_n$ and $m = \lceil c_{\mathcal{H}}\, n \rceil$ with $c_{\mathcal{H}}$ as in Theorem 2.8. Then,*

$$\Pr\left[\operatorname{adeg}(k;\, \mathsf{M}_m) \in (1 \pm \varepsilon)\, a_k\, n \,\middle|\, C(\mathsf{M}_m) = C\right] \ge 1 - e^{-\varepsilon^2 n/\alpha(n)},$$

where M_m is drawn uniformly at random from $\mathcal{M}_m$ and $a_k = [z^k]\big\{E_{\mathcal{H}}(z) - 2\, R_{\mathcal{H}}(z)\big\}$.

Proof. Let us first consider the case $a_k = 0$. Then, $[z^k]\{E_{\mathcal{H}}(z) - 2\, R_{\mathcal{H}}(z)\} = 0$ holds, which implies that not a single graph in $\mathcal{H}$ has a vertex different from the endpoints of the root edge with degree equal to k. Hence, as $M = C \circ (H_1, \ldots, H_{|C|})$, we have $\operatorname{adeg}(k;\, \mathsf{M}_m) = 0$ with probability 1.

Suppose in the remainder that $a_k > 0$, and set $\beta(n) = \min\{\alpha(n)^{1/4}, \log n\}$. Let M be a random map from $\mathcal{M}$, drawn according to the Boltzmann distribution $\Gamma M(\rho_{\mathcal{M}})$. Consider the two events

$$\begin{aligned} &(\mathfrak{A}) \quad \operatorname{adeg}(k;\, \mathsf{M}) \notin (1 \pm \varepsilon)\, a_k\, n \quad \text{and} \\ &(\mathfrak{B}) \quad \mathsf{M} \in \mathcal{M}_m \wedge C(\mathsf{M}) = C. \end{aligned}$$

Since in the Boltzmann distribution $\Gamma M(\rho_M)$ all maps with a given number of edges have the same probability of being M, we obtain

$$\Pr\left[\operatorname{adeg}(k;\, \mathsf{M}_m) \notin (1 \pm \varepsilon)\, a_k\, n \,\middle|\, C(\mathsf{M}_m) = C\right] = \Pr\left[\mathfrak{A} \,\middle|\, \mathfrak{B}\right]. \tag{2.3.8}$$

Suppose that $C(\mathsf{M}) = C$ and recall that $|C| = n$. By Lemma 2.11 and the fact that $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ is a proper composition schema, M decomposes into its core C and exactly n substitution maps $\mathsf{H}^{(1)}, \ldots, \mathsf{H}^{(n)}$ drawn independently from $\mathcal{H}$ according to the Boltzmann distribution $\Gamma H(\rho_{\mathcal{H}})$. Thus, (2.3.8) is equivalent to

$$\Pr[\mathfrak{A} \mid \mathfrak{B}] = \Pr\left[\sum_{i=1}^{n} \deg^*(k;\, \mathsf{H}^{(i)}) \notin (1 \pm \varepsilon)\, a_k\, n \,\middle|\, \mathfrak{B}\right] \tag{2.3.9}$$

where $\deg^*(k;\, H)$ counts in $H \in \mathcal{H}$ the number of vertices of degree k distinct from the endpoints of the root edge. Let

$$\begin{aligned} X &:= \sum_{i=1}^{n} |\mathsf{H}^{(i)}| \cdot \chi_{\{|\mathsf{H}^{(i)}| \le \beta(n)\}} \quad \text{and} \\ Y &:= \sum_{i=1}^{n} \deg^*(k;\, \mathsf{H}^{(i)}) \cdot \chi_{\{|\mathsf{H}^{(i)}| \le \beta(n)\}}, \end{aligned}$$

where $\chi_{\mathfrak{G}} \in \{0, 1\}$ is the indicator function of the event $\mathfrak{G}$, that is, $\chi_{\mathfrak{G}}$ is one if the event occurs and zero otherwise. We consider the two events

$$\begin{aligned} &(\mathfrak{E}) \quad X \le \left(1 - \frac{\varepsilon\, a_k\, n}{2\, m}\right) m \quad \text{and} \\ &(\mathfrak{F}) \quad Y \notin \left(1 \pm \frac{\varepsilon}{2}\right) a_k\, n\,. \end{aligned}$$

Suppose $\mathfrak{B}$ holds. We show that in this case $\neg\mathfrak{E}$ and $\neg\mathfrak{F}$ together imply $\neg\mathfrak{A}$ and thus (2.3.9) infers that

$$\Pr[\mathfrak{A} \mid \mathfrak{B}] \le \Pr[\mathfrak{E} \vee \mathfrak{F} \mid \mathfrak{B}]\,. \tag{2.3.10}$$

Indeed, suppose $\mathsf{M} \in \mathcal{M}_m$. Then $m = \sum_{i=1}^{n} |\mathsf{H}^{(i)}|$ and $\neg\mathfrak{E}$ implies

$$\sum_{i=1}^{n} |\mathsf{H}^{(i)}| \cdot \chi_{\{|\mathsf{H}^{(i)}| > \beta(n)\}} \le \frac{\varepsilon}{2}\, a_k\, n\,.$$

Since $\deg^*(k;\, H) \le |H|$ for all $H \in \mathcal{H}$ (these maps are connected and the end-vertices of the root ignored), also

$$\sum_{i=1}^{n} \deg^*(k;\, \mathsf{H}^{(i)}) \cdot \chi_{\{|\mathsf{H}^{(i)}| > \beta(n)\}} \le \frac{\varepsilon}{2}\, a_k\, n\,.$$

Thus, together with $\neg\mathfrak{F}$ this implies $\neg\mathfrak{A}$ and thus (2.3.10). Elementary probability theory now gives us

$$\Pr[\mathfrak{A} \mid \mathfrak{B}] \le \frac{\Pr[\mathfrak{E}] + \Pr[\mathfrak{F}]}{\Pr[\mathfrak{B}]} \tag{2.3.11}$$

and finally frees us from the condition $\mathfrak{B}$. Applying Theorem 2.8 and Corollary 2.10 yields

$$\Pr[\mathfrak{B}] = \Pr[C(\mathsf{M}_m) \in \mathcal{C}_n \mid \mathsf{M} \in \mathcal{M}_m] \cdot \Pr[\mathsf{M} \in \mathcal{M}_m] = \Theta(n^{-19/6}). \tag{2.3.12}$$

Moreover, again by Corollary 2.10, if we denote by H a map drawn from $\mathcal{H}$ according to the Boltzmann distribution $\Gamma H(\rho_{\mathcal{H}})$, then

$$\mathrm{E}\Big[\sum_{i=1}^{n} |\mathsf{H}^{(i)}| \cdot \chi_{\{|\mathsf{H}^{(i)}|>\beta(n)\}}\Big] = \sum_{\ell>\beta(n)} \ell \cdot \Pr\big[|\mathsf{H}| = \ell\big]\, n = \mathcal{O}(n^{1/2})\,.$$

Since, $\deg^*(k;\, H) \leq |H|$ for any $H \in \mathcal{H}$, also

$$\mathrm{E}\Big[\sum_{i=1}^{n} \deg^*(k;\, \mathsf{H}^{(i)}) \cdot \chi_{\{|\mathsf{H}^{(i)}|>\beta(n)\}}\Big] = \mathcal{O}(n^{1/2}).$$

Thus, $\mathrm{E}[X] = m + o(n)$ and $\mathrm{E}[Y] = a_k\, n + o(n)$ holds by the definition of a_k.

In order to show that with high probability X and Y do not deviate much from their expectations we apply the inequality by Azuma and Hoeffding (see Theorem 4.3 in Part II, Chapter 4). For $i \in \{1, \ldots, n\}$, the choice of $\mathsf{H}^{(i)}$ can change the values of X or Y by at most $\beta(n)$. Thus, by the Azuma-Hoeffding inequality,

$$\Pr[\,\mathfrak{E}\,] \leq e^{-\varepsilon^2 n/\beta(n)^3} \quad \text{and}$$
$$\Pr[\,\mathfrak{F}\,] \leq e^{-\varepsilon^2 n/\beta(n)^3}\,.$$

The proof completes by combining (2.3.11), (2.3.12), the two previous inequalities and the fact that $\beta(n)^4 \leq \alpha(n)$. ☐

The next proposition asserts that also (2.3.4) holds with high probability under the condition $C(\mathsf{M}_m) \in \mathcal{C}_n$.

Proposition 2.14. *Let $\mathcal{M} = \mathcal{C} \circ \mathcal{H} + \mathcal{D}$ be a proper map composition schema. Let $\varepsilon > 0$, $\gamma > 0$, $k, \ell \in \mathbb{N}$, $\alpha \in \omega(1)$, and $n \in \mathbb{N}$ sufficiently large. Furthermore, let $C \in \mathcal{C}_n$ with $\deg(\ell;\, C) \geq \gamma\, n$, and $m = \lceil c_{\mathcal{H}}\, n \rceil$ with $c_{\mathcal{H}}$ as in Theorem 2.8. Then,*

$$\Pr\big[\mathrm{bdeg}(k, \ell;\, \mathsf{M}_m) \in (1 \pm \varepsilon)\, b_{k,\ell} \deg(\ell;\, C) \,\big|\, C(\mathsf{M}_m) = C\big] \geq 1 - e^{-\varepsilon^2 n/\alpha(n)},$$

where M_m is a map drawn uniformly at random from $\mathcal{M}_m$ and $b_{k,\ell} = [z^k] R_{\mathcal{H}}(z)^\ell$.

Proof. Suppose that $b_{k,\ell} > 0$ (otherwise, the statement holds trivially). Let $n \in \mathbb{N}$ be sufficiently large, and let $\beta(n) = \alpha(n)^{1/2}$. Let M be a random map from $\mathcal{M}$, drawn according to the Boltzmann distribution $\Gamma M(\rho_{\mathcal{M}})$. Consider the two events

$$(\mathfrak{A}) \quad \mathrm{bdeg}(k, \ell;\, \mathsf{M}) \notin (1 \pm \varepsilon)\, b_{k,\ell} \deg\big(\ell;\, C\big) \quad \text{and}$$
$$(\mathfrak{B}) \quad \mathsf{M} \in \mathcal{M}_m \wedge C(\mathsf{M}) = C.$$

Since in the Boltzmann distribution all maps with a given number of edges have the same probability of being M, we obtain

$$\Pr[\mathrm{bdeg}(k, \ell;\, \mathsf{M}_m) \in (1 \pm \varepsilon)\, b_{k,\ell} \deg\big(\ell;\, C\big) \,\big|\, C(\mathsf{M}_m) = C] = \Pr[\,\mathfrak{A} \mid \mathfrak{B}]\,. \tag{2.3.13}$$

Elementary probability theory then implies

$$\Pr[\mathfrak{A} \mid \mathfrak{B}] \leq \frac{\Pr[\mathfrak{A} \,\big|\, C(\mathsf{M}) = C]}{\Pr[\mathsf{M} \in \mathcal{M}_m \mid C(\mathsf{M}) = C]} \tag{2.3.14}$$

Suppose $C(\mathsf{M}) = C$. Then by Lemma 2.11, $\mathsf{M} = C \circ (\mathsf{H}^{(1)}, \ldots \mathsf{H}^{(n)})$ where the maps $(\mathsf{H}^{(i)})_{1 \leq i \leq n}$ are drawn mutually independent from $\mathcal{H}$ according to the Boltzmann distribution $\Gamma H(\rho_{\mathcal{H}})$. Furthermore

$$\mathrm{bdeg}(k, \ell; \mathsf{M}) = \sum_{v \in C} \chi_{\{\deg(v,C)=\ell \,\wedge\, \deg(v,\mathsf{M})=k\}} \tag{2.3.15}$$

where $\chi_{\mathfrak{G}} \in \{0, 1\}$ is the indicator function of the event $\mathfrak{G}$.

Next, let $v \in C$ such that $\deg(v; C) = \ell$ and let $\mathsf{H}^{(\sigma_1)}, \ldots, \mathsf{H}^{(\sigma_\ell)}$ be the substitution maps from $\mathcal{H}$ containing v (that is, the edges incident with v in C were replaced by precisely those $\mathsf{H}^{(i)}$'s). Then, by the symmetry of the $\mathsf{H}^{(i)}$'s with respect to the root,

$$\deg(v, \mathsf{M}) = \sum_{j=1}^{\ell} \mathrm{rdeg}(\mathsf{H}^{(\sigma_j)}) \,.$$

Thus, since $b_{k,\ell} = [z^k] R_{\mathcal{H}}(z)^\ell$ is the probability that the sum of the root degrees of ℓ maps drawn mutually independently from $\mathcal{H}$ according to the Boltzmann distribution $\Gamma H(\rho_{\mathcal{H}})$ equals k, we obtain

$$\mathrm{E}\big[\chi_{\{\deg(v,C)=\ell \,\wedge\, \deg(v,\mathsf{M})=k\}}\big] = b_{k,\ell} \cdot \chi_{\{\deg(v,C)=\ell\}} \,.$$

Thus, by (2.3.15),

$$\mathrm{E}[\mathrm{bdeg}(k, \ell; \mathsf{M}) \mid C(\mathsf{M}) = C] = b_{k,\ell} \deg(\ell;\, C) \,.$$

For $i \in \{1, \ldots, n\}$, the influence of the single substitution map $\mathsf{H}^{(i)}$ on $\mathrm{bdeg}(k, \ell; \mathsf{M})$ is at most two, as the endpoints of the root edge of $\mathsf{H}^{(i)}$ are identified with the endpoint of some edge in C, and no other vertices of C are affected. Thus, if we condition on $C(\mathsf{M}) = C$, then for sufficiently large n the Azuma-Hoeffding bounds (see Theorem 4.3 in Part II, Chapter 4) yield

$$\Pr[\mathrm{bdeg}(k, \ell; \mathsf{M}) \notin (1 \pm \varepsilon)\, b_{k,\ell} \deg(\ell;\, C) \,\big|\, C(\mathsf{M}) = C] \leq e^{-\varepsilon^2 n/\beta(n)} \,.$$

Moreover, $\Pr[M \in \mathcal{M}_m \mid C(M) = C] = \Theta(n^{-2/3})$ holds by Corollary 2.10. Then, the claimed statement follows from (2.3.14), the previous inequality, and the definition of $\beta(n)$. □

With the previous two propositions at hand we are finally able to prove Theorem 2.12.

Proof of Theorem 2.12. Let for brevity $d_{\mathcal{M},\ell} = [z^\ell] D_{\mathcal{M}}(z)$ and $d_{\mathcal{C},\ell} = [z^\ell] D_{\mathcal{C}}(z)$. We show by induction on k that for $\varepsilon \in (0, 1)$, $\alpha(n) \in \omega(1)$ and sufficiently large n

$$\Pr[\deg(k; \mathsf{C}_n) \in (1 \pm \varepsilon)\, d_{\mathcal{C},k}\, n] \geq 1 - h_\alpha(\varepsilon, n).$$

For $k = 0$ this statement holds trivially since $\deg(0; \mathsf{C}_n) = 0$. Thus, let $k \geq 1$, $\varepsilon \in (0,1)$, $\alpha(n) \in \omega(1)$ and n sufficiently large. Let $\beta(n) = \alpha(n)^{1/3}$, and $m = \lceil c_{\mathcal{H}}\, n \rceil$. By the induction hypothesis, for sufficiently large n we may assume that

$$\forall\, 0 \leq \ell < k\colon \quad \Pr[\deg(\ell; \mathsf{C}_n) \in (1 \pm \varepsilon/6)\, d_{\mathcal{C},\ell}\, n] \geq 1 - h_\beta(\varepsilon/6, n)\,. \tag{2.3.16}$$

Let M_m be drawn uniformly at random from $\mathcal{M}_m$ and let $C = C(\mathsf{M}_m)$. If $C \in \mathcal{C}_n$, then C is distributed uniformly and thus

$$\Pr[\deg(k; \mathsf{C}_n) \in (1 \pm \varepsilon)\, d_{\mathcal{C},k}\, n] = \Pr[\deg(k;\, C) \in (1 \pm \varepsilon)\, d_{\mathcal{C},k}\, n \mid C \in \mathcal{C}_n]\,. \tag{2.3.17}$$

Consider the two events

$$\begin{aligned} &(\mathfrak{A}) \quad \mathrm{bdeg}(k,k;\, \mathsf{M}_m) \in (1 \pm \varepsilon/2)\, b_{k,k} \deg(k;\, C) \quad \text{and} \\ &(\mathfrak{B}) \quad \mathrm{bdeg}(k,k;\, \mathsf{M}_m) \in (1 \pm \varepsilon/2)\, b_{k,k}\, d_{\mathcal{C},k}\, n\,. \end{aligned}$$

If $\mathfrak{A}$ and $\mathfrak{B}$ hold simultaneously, then also $\deg(k;\, C) \in (1 \pm \varepsilon)\, d_{\mathcal{C},k}\, n$. Thus,

$$\Pr[\deg(k;\, C) \in (1 \pm \varepsilon)\, d_{\mathcal{C},k}\, n \mid C \in \mathcal{C}_n] \geq \Pr[\mathfrak{A} \mid \mathfrak{B} \wedge C \in \mathcal{C}_n] \cdot \Pr[\mathfrak{B} \mid C \in \mathcal{C}_n]\,. \tag{2.3.18}$$

Suppose $\mathfrak{B}$ holds. Then, $\deg(k;\, C) \geq \mathrm{bdeg}(k,k; \mathsf{M}_m) \geq \gamma\, n$ for some $\gamma > 0$. Thus, we can apply Proposition 2.14 to show that

$$\Pr[\mathfrak{A} \mid \mathfrak{B} \wedge C(\mathsf{M}_m) \in \mathcal{C}_n] \geq 1 - h_{\beta^2}(\varepsilon, n)\,.$$

Thus, if we also show

$$\Pr[\mathfrak{B} \mid C(\mathsf{M}_m) \in \mathcal{C}_n] \geq 1 - (2k+2)\, h_{\beta^2}(\varepsilon, n)\,. \tag{2.3.19}$$

then Theorem 2.12 follows for sufficiently large n from (2.3.17), (2.3.18) and the fact that $\beta(n) = \alpha(n)^{1/3}$. Recall that by (2.3.3)

$$\mathrm{bdeg}(k,k;\, \mathsf{M}_m) = \deg(k;\, \mathsf{M}_m) - \mathrm{adeg}(k;\, \mathsf{M}_m) - \sum_{\ell=0}^{k-1} \mathrm{bdeg}(k,\ell;\, \mathsf{M}_m) \tag{2.3.20}$$

and that by (2.3.7)

$$b_{k,k}\, d_{\mathcal{C},k} = d_{\mathcal{M},k}\, c_{\mathcal{H}} - a_k - \sum_{\ell=0}^{k-1} b_{k,\ell}\, d_{\mathcal{C},\ell} \tag{2.3.21}$$

with $a_k = [z^k]\{E_{\mathcal{H}}(z) - 2\, R_{\mathcal{H}}(z)\}$ and $b_{k,\ell} = [z^k]\{R_{\mathcal{H}}(z)^\ell\}$ for all $k, \ell \in \mathbb{N}$. We show that with high probability each quantity on the right-hand side of (2.3.20) is concentrated around n times the corresponding quantity in (2.3.21).

Since n is sufficiently large, (2.3.6) together with Theorem 2.8 implies by elementary probability theory that

$$\Pr[\deg(k;\, \mathsf{M}_m) \in (1 \pm \varepsilon/2)\, d_{\mathcal{M},k}\, c_{\mathcal{H}}\, n \mid C(\mathsf{M}_m) \in \mathcal{C}_n] \geq 1 - h_\beta(\varepsilon, n)\,. \tag{2.3.22}$$

since n is sufficiently large, $m \geq n$, and g is monotone in both arguments. Next, by Proposition 2.13 and n sufficiently large,

$$\Pr[\mathrm{adeg}(k;\, \mathsf{M}_m) \in (1 \pm \varepsilon/2)\, a_k\, n \mid C(\mathsf{M}_m) \in \mathcal{C}_n] \geq 1 - h_{\beta^2}(\varepsilon, n). \tag{2.3.23}$$

Finally, by Proposition 2.14 and the induction hypothesis (2.3.16), and again n sufficiently large,

$$\Pr[\mathrm{bdeg}(k, \ell;\, \mathsf{M}_m) \in (1 \pm \varepsilon/2)\, b_{k,\ell}\, n \mid C(\mathsf{M}_m) \in \mathcal{C}_n] \geq 1 - h_{\beta}(\varepsilon, n) - h_{\beta^2}(\varepsilon, n) \tag{2.3.24}$$

for all $0 \leq \ell < k$. Hence, for sufficiently large n, (2.3.19) holds by (2.3.20) and (2.3.21), together with (2.3.22)–(2.3.24). This concludes the proof of Theorem 2.12. □

2.4. Biconnected Maps

In this section, we apply the framework from the previous section. Our main strategy is to prove that the preconditions of Theorem 2.12 are satisfied. First, we show that the class $\mathcal{G}$ of general maps, and the class $\mathcal{B}$ of biconnected maps can be related through a proper map composition schema.

Lemma 2.15 (Tutte (1963)). *Let "$\bullet\!\!\longrightarrow\!\!\bullet$" be the map that consists of two vertices and a single edge, and "$\bullet$" the empty map that consists of a single vertex (and no edge). The classes of general maps $\mathcal{G}$ and biconnected maps $\mathcal{B}$ satisfy the relation*

$$\mathcal{G} = \mathcal{B} \circ \mathcal{H}\,,$$

where

$$\mathcal{H} = \{\bullet\!\!\longrightarrow\!\!\bullet\} \times (\mathcal{G} + \{\bullet\})^2\,.$$

Combinatorially, $\mathcal{H}$ is the class of general maps where the root edge is a bridge, that is, every $H \in \mathcal{H}$ consists of two submaps $G_1, G_2 \in \mathcal{G} + \{\bullet\}$ whose root vertices are joined through an additional edge that is distinguished as the root edge of the composed map (where G_1 or G_2 may be the empty map "$\bullet$"). See Figure 2.2 for an example.

We see that $\mathcal{H}$ is closed under inversion of the orientation of the root edge, and moreover, Table 2.1 guarantees that the generating functions $G(x)$ and $B(x)$

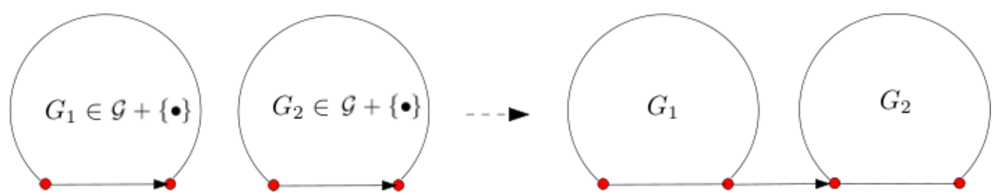

Figure 2.2. Composing maps in $\mathcal{H}$ out of maps in $\mathcal{G}$.

are singular with exponent 3/2. Finally, a straightforward calculation shows that also $H(x) = x(G(x)+1)^2$ is singular with exponent 3/2, admitting the expansion

$$H(x) = \tfrac{4}{27} - \tfrac{4}{9}\left(1 - \tfrac{x}{\rho_{\mathcal{G}}}\right) + \tfrac{16}{27}\left(1 - \tfrac{x}{\rho_{\mathcal{G}}}\right)^{3/2} + \mathcal{O}\Big(\left(1 - \tfrac{x}{\rho_{\mathcal{G}}}\right)^2\Big) \tag{2.4.1}$$

as $x \to \rho_{\mathcal{H}} = \rho_{\mathcal{G}} = \frac{1}{12}$.

The above facts together with Theorem 2.1 imply that the preconditions of Theorem 2.12 are satisfied. So, we obtain Theorem 2.2, where the function $D_{\mathcal{B}}(z)$ is given by the relation

$$3\,D_{\mathcal{G}}(z) = E_{\mathcal{H}}(z) - 2R_{\mathcal{H}}(z) + D_{\mathcal{B}}\big(R_{\mathcal{H}}(z)\big), \tag{2.4.2}$$

where $D_{\mathcal{G}}(z)$ is given in Theorem 2.1. Moreover, the next lemma gives explicit expressions for $R_{\mathcal{H}}(z)$ and $E_{\mathcal{H}}(z)$, and derives an asymptotic expression for $[z^k]D_{\mathcal{B}}(z)$.

Lemma 2.16. $[z^k]D_{\mathcal{B}}(z) \sim_k \sqrt{\frac{9}{2\pi}}k^{-1/2}\left(\frac{2}{3}\right)^k$. *Moreover, $R_{\mathcal{H}}(z)$ and $E_{\mathcal{H}}(z)$ are given explicitly by*

$$R_{\mathcal{H}}(z) = \frac{3z^2 - 36z + 36 - \sqrt{3\,(z+2)\,(6-5z)^3}}{8z\,(1-z)} \quad \textit{and}$$
$$E_{\mathcal{H}}(z) = \frac{18 - 3z - \sqrt{9\,(z+2)\,(6-5z)}}{2z}.$$

Let us make two auxiliary preparations before we actually prove the lemma. In the case of $\mathcal{G}$ the ordinary generating function $G(x)$ can be determined explicitly, see e.g. Banderier et al. (2001). It is given by

$$G(x) = \frac{-1 + 18x + (1-12x)^{3/2}}{54x^2} - 1. \tag{2.4.3}$$

Moreover, the bivariate function $G(x,z)$, where x marks the size and z the root face degree of a map is defined through

$$G(x,z) = xz^2\,(G(x,z)+1)^2 + xz\,\frac{(G(x)+1) - z\,(G(x,z)+1)}{1-z}. \tag{2.4.4}$$

Proof of Lemma 2.16. First of all, note that as the root degree of a random map from $\mathcal{H}$ is one plus the root degree of a random map from $\mathcal{G} + \{\bullet\}$. Thus, we obtain by Proposition 2.9,

$$R_{\mathcal{H}}(z) = z\,\frac{G(\rho_{\mathcal{G}}, z) + 1}{G(\rho_{\mathcal{G}}) + 1}.$$

By using the explicit expressions for $G(x)$ and $G(x,z)$ from (2.4.3) and (2.4.4) and the fact $\rho_{\mathcal{G}} = \frac{1}{12}$ from Table 2.1 we arrive at the explicit expression for $R_{\mathcal{H}}(z)$ in Lemma 2.16.

In order to obtain $E_{\mathcal{H}}(z)$ we first determine $E_{\mathcal{G}}(z)$. By applying again Proposition 2.9, this time the third statement, we obtain an explicit expression for $E_{\mathcal{G}}(z)$.

Then, by linearity of expectation $E_{\mathcal{H}}(z) = 2(E_{\mathcal{G}+\{\bullet\}}(z) - R_{\mathcal{G}+\{\bullet\}}(z)) + 2R_{\mathcal{H}}(z)$, which implies the explicit expression for $E_{\mathcal{H}}(z)$ in Lemma 2.16.

To obtain the asymptotic form of the coefficients first note that the function $R_{\mathcal{H}}(z)$ is strictly increasing for $z \in [0, \frac{6}{5}]$, and that $R_{\mathcal{H}}(\frac{6}{5}) = \frac{3}{2}$. Hence, $R_{\mathcal{H}}$ is uniquely invertible in $[0, \frac{6}{5}]$, the inverse being

$$F(z) = \frac{27 + 36z + 4z^2 - \sqrt{729 - 1512z + 1080z^2 - 288z^3 + 16z^4}}{2(24 + 3z + 4z^2)}\,.$$

We thus obtain

$$D_{\mathcal{B}}(z) = 3D_{\mathcal{G}}(F(z)) - E_{\mathcal{H}}(F(z)) + 2R_{\mathcal{H}}(F(z)).$$

Now, as $D_{\mathcal{G}}, E_{\mathcal{H}}, R_{\mathcal{H}}$, and F are given explicitly, we obtain an explicit expression for $D_{\mathcal{B}}(z)$. We see that the dominant singularity of $D_{\mathcal{B}}(z)$ is at $\frac{3}{2}$, and that

$$D_{\mathcal{B}}(z) = \sqrt{\tfrac{9}{2}}\left(1 - \tfrac{2}{3}\,z\right)^{-1/2} + \mathcal{O}(1)$$

as $z \to \frac{3}{2}$.

The proof finishes by applying the Transfer Theorem (Theorem 2.5) to the above local expansion. $\square$

2.5. c-Nets

This section deals with the proof of Theorem 2.3. Again, we use the framework from Section 2.3 and show that the preconditions of Theorem 2.12 are satisfied. The proofs in this section follows the same lines as those of the previous section, but the details are considerably more involved. In particular, we now consider classes of maps where the root is not marked.

Let $\mathcal{F}$ be a class of maps where all edges (including the root edge) are marked. Let $\bullet\!\circlearrowleft$ be the map that contains a single loop, and $\bullet\!\!\rightarrow\!\!\bullet$ be the map containing a single edge. We define the *unmarked* class $\mathcal{F}^\circ$ by removing $\bullet\!\circlearrowleft$ and $\bullet\!\!\rightarrow\!\!\bullet$ from $\mathcal{F}$, and by unmarking the root-edge of any other graph that remains in $\mathcal{F}$. More formally, $\mathcal{F}$ and $\mathcal{F}^\circ$ are related through

$$\mathcal{F} = \mathcal{F}^\circ \times \{\bullet\!\!\rightarrow\!\!\bullet\} + \{\bullet\!\circlearrowleft\} + \{\bullet\!\!\rightarrow\!\!\bullet\}\,,$$

where "$\mathcal{F}^\circ \times \mathcal{X}$" means that we identify the roots of maps from $\mathcal{F}^\circ$ and $\mathcal{X}$, resulting in a map with a marked root. It follows for the corresponding generating functions that

$$F(x) = x\,F^\circ(x) + 2x \quad \text{and}$$
$$F(x,z) = xz\,F^\circ(x,z) + xz^2 + xz\,.$$

With all the above notation at hand we can describe the composition of biconnected maps by c-nets. First of all, let $\mathcal{B}$ be the class of biconnected maps, where by convention we assume that $\bullet\!\circlearrowleft, \bullet\!\!\rightarrow\!\!\bullet \in \mathcal{B}$. The *core* $C(B)$ of a biconnected map $B \in \mathcal{B}$ is obtained by cutting all maximal 2-separators and by replacing the removed components by edges. It can be shown that the core of a biconnected map is either a c-net or empty. More precisely, Tutte (1963) showed the following combinatorial relation.

Lemma 2.17 (Tutte (1963)). *Let $\mathcal{D} \subset \mathcal{B}$ be the class of biconnected maps that have an empty 3-connected core. The classes $\mathcal{B}$ and $\mathcal{T}$ satisfy the relation*

$$\mathcal{B}^\circ = \mathcal{T}^\circ \circ \mathcal{B}^\circ + \mathcal{D}^\circ .$$

In words, any map in $\mathcal{B}^\circ$ has either an empty core, or it is obtained from a c-net by substituting every non-root edge by a map in $\mathcal{B}^\circ$, and then removing that roots of the substitution maps (see also Section 2.1).

In order to apply Theorem 2.12 we develop further the composition schema that is described in Lemma 2.17. More specifically, our aim is to find an explicit relation between the class of c-nets and the class of general maps $\mathcal{G}$, for which we have sufficiently strong concentration results for the number of vertices of degree k, see Theorem 2.1. To this purpose we reconsider the composition scheme from Lemma 2.15, but this time applied to maps from $\mathcal{B}^\circ$, that is, biconnected maps with unmarked roots. Let $\mathcal{H}$ be defined as in Lemma 2.15, see also Figure 2.2. Define the class of maps $\mathcal{G}^*$ by

$$\mathcal{G}^* = \mathcal{B}^\circ \circ \mathcal{H} .$$

Note that the root edges of the maps in $\mathcal{G}^*$ are not marked, and moreover that $\mathcal{G}^*$ is a proper subset of $\mathcal{G}^\circ$, that is, $\mathcal{G}^* \neq \mathcal{G}^\circ$. In particular, in order to generate all maps in $\mathcal{G}$ (except for $\bullet\!\circlearrowleft$ and $\bullet\!\!\rightarrow\!\!\bullet$), we need to substitute the roots from the maps in $\mathcal{G}^*$ by maps in $\mathcal{H}$ and not by edges. (this can be seen from Lemma 2.15). Thus,

$$\mathcal{G} = \mathcal{G}^* \times \mathcal{H} + \{\bullet\!\circlearrowleft\} \circ \mathcal{H} + \{\bullet\!\!\rightarrow\!\!\bullet\} \circ \mathcal{H} \tag{2.5.1}$$

and if we substitute $H(x) = x\,(G(x)+1)^2$ and $H(x,z) = x\,z\,(G(x)+1)\,(G(x,z)+1)$ in the corresponding generating functions, we obtain (with extra care for the term corresponding to $\{\bullet\!\circlearrowleft\} \circ \mathcal{H}$),

$$G^*(x) = \frac{G(x)}{x\,(G(x)+1)^2} - 2 \quad \text{and} \tag{2.5.2}$$

$$G^*(x,z) = \frac{G(x) - xz^2\,(G(x,z)+1)^2}{x\,z\,(G(x)+1)\,(G(x,z)+1)} - 1 . \tag{2.5.3}$$

Hence, we can transfer results for $\mathcal{G}$ to $\mathcal{G}^*$. Moreover, we can relate $\mathcal{G}^*$ to $\mathcal{T}^\circ$ by substituting the edges of all maps from the composition schema in Lemma 2.17 by maps from $\mathcal{H}$ and then applying the definition of $\mathcal{G}^*$.

Lemma 2.18. *Let $\mathcal{D} \subset \mathcal{B}$ be the class of biconnected maps that have an empty 3-connected core. The classes $\mathcal{G}^*$ and $\mathcal{T}$ satisfy the relation*

$$\mathcal{G}^* = \mathcal{T}^\circ \circ \mathcal{G}^* + \mathcal{D}^\circ \circ \mathcal{H} .$$

With the previous lemma at hand, our proof strategy for Theorem 2.3 is as follows. We will check that the preconditions of Theorem 2.12 are satisfied. In particular, we show (i) that the degree sequence of a map G^*_n chosen uniformly at random from $\mathcal{G}^*_n$ is asymptotically the same as that of a map G_n chosen uniformly at random from the class of general maps $\mathcal{G}_n$; (ii) we assert that $\mathcal{G}^* = \mathcal{T}^\circ \circ \mathcal{G}^* + \mathcal{D}^\circ \circ \mathcal{H}$ is a proper map composition schema; and (iii) we determine the generating function $D_{\mathcal{T}}(z)$ and its asymptotic behavior.

We first carry out step (i).

Proposition 2.19. *Let $\varepsilon > 0$, $k \in \mathbb{N}$. Let G_n^* be a map drawn uniformly at random from $\mathcal{G}_n^*$. Then, uniformly for large n*

$$\Pr\left[\deg(k;\, \mathsf{G}_n^*) \in (1 \pm \varepsilon)\, g_k\, n\right] \geq 1 - \mathcal{O}\big(\varepsilon^{-2}(g_k\, n)^{-1} \log^{20} n\big)\,, \tag{2.5.4}$$

where $g_k = [z^k]D_{\mathcal{G}}(z)$ and $D_{\mathcal{G}}(z)$ as in Theorem 2.1.

Proof. Let $k \in \mathbb{N}$ be fixed and $n \in \mathbb{N}$ be sufficiently large. Let G be drawn from $\mathcal{G}$ according to the Boltzmann distribution $\Gamma G(\rho_G)$, let G_n^* be drawn uniformly at random from $\mathcal{G}_n^*$, land G_n be drawn uniformly at random from $\mathcal{G}_n$. Then

$$\begin{aligned}
\Pr\left[\deg(k; \mathsf{G}_n) \notin (1 \pm \varepsilon)\, g_k\, n\right] &= \Pr\left[\deg(k; \mathsf{G}) \notin (1 \pm \varepsilon)\, g_k\, n \,\middle|\, \mathsf{G} \in \mathcal{G}_n\right] \quad \text{and} \\
\Pr\left[\deg(k; \mathsf{G}_n^*) \notin (1 \pm \varepsilon)\, g_k\, n\right] &= \Pr\left[\deg(k; \mathsf{G}) \notin (1 \pm \varepsilon)\, g_k\, n \,\middle|\, \mathsf{G} \in \mathcal{G}_n^* \times \{\bullet\!\!\rightarrow\!\!\bullet\}\right],
\end{aligned}$$

where $\mathsf{G} \in \mathcal{G}_n^* \times \{\bullet\!\!\rightarrow\!\!\bullet\}$ is defined as in (2.5.1). Whenever $\mathsf{G} \in \mathcal{G}_n^* \times \{\bullet\!\!\rightarrow\!\!\bullet\}$, then certainly $\mathsf{G} \in \mathcal{G}_n$. Thus,

$$\Pr\left[\deg(k; \mathsf{G}_n^*) \notin (1 \pm \varepsilon)\, g_k\, n\right] \leq \frac{\Pr[\mathsf{G} \in \mathcal{G}_n]}{\Pr\left[\mathsf{G} \in \mathcal{G}_n^* \times \{\bullet\!\!\rightarrow\!\!\bullet\}\right]} \cdot \Pr\left[\deg(k; \mathsf{G}_n) \notin (1 \pm \varepsilon)\, g_k\, n\right].$$

Now, $\Pr[\mathsf{G} \in \mathcal{G}_n] = \Theta(n^{-5/2})$ by Corollary 2.10. Moreover, by exploiting (2.5.2) and (2.4.3) we can derive the singular expansion of $G^*(x) = \frac{G(x)}{H(x)} - 2$, which has the expansion

$$G^*(x) = \frac{1}{4} - \frac{9}{4}\left(1 - \tfrac{x}{\rho_{\mathcal{G}}^*}\right) + 9\left(1 - \tfrac{x}{\rho_{\mathcal{G}}^*}\right)^{3/2} + \mathcal{O}\Big(\big(1 - \tfrac{x}{\rho_{\mathcal{G}}^*}\big)^2\Big) \tag{2.5.5}$$

as $x \to \rho_{\mathcal{G}}^* = \rho_{\mathcal{G}} = \rho_{\mathcal{H}} = \frac{1}{12}$.

Hence, $G^*(x)$ is singular with exponent $3/2$, and an application of Theorem 2.6 yields

$$\Pr\left[\mathsf{G} \in \mathcal{G}_n^* \times \{\bullet\!\!\rightarrow\!\!\bullet\}\right] = \frac{[x^n]xG^*(x) \cdot \rho_{\mathcal{G}}^n}{G(\rho_{\mathcal{G}})} = \Theta(n^{-5/2})\,.$$

The statement follows from Theorem 2.1. □

Next, we show that $\mathcal{G}^* = \mathcal{T}^\circ \circ \mathcal{G}^* + \mathcal{D}^\circ \circ \mathcal{H}$ is a proper map composition schema. It can be seen that since $\mathcal{G}$ and $\mathcal{H}$ are closed under inversion of the orientation of the root, so is $\mathcal{G}^*$. Table 2.1 guarantees that the generating functions $M(x)$, $H(x)$, and $T^\circ(x)$ are singular with exponent $3/2$. Moreover, in (2.5.5) we showed also that $G^*(x)$ is singular with exponent $3/2$.

The above facts together with Proposition 2.19 imply that the preconditions of Theorem 2.12 are satisfied. We obtain Theorem 2.3, where $D_{\mathcal{T}}(z) = D_{\mathcal{T}^\circ}(z)$ is given by the relation

$$9\, D_{\mathcal{G}^*}(z) = E_{\mathcal{G}^*}(z) - 2R_{\mathcal{G}^*}(z) + D_{\mathcal{T}}\big(R_{\mathcal{G}^*}(z)\big), \tag{2.5.6}$$

where $D_{\mathcal{G}*}(z) = D_{\mathcal{G}}(z)$ is given in Theorem 2.1 and $c_{\mathcal{G}^*} = \frac{-9/4}{1/4} = 9$. Moreover, the proof of the next lemma derives explicit expressions for $R_{\mathcal{G}^*}(z)$ and $E_{\mathcal{G}^*}(z)$, albeit the second is omitted.

Lemma 2.20. $[z^k]D_{\mathcal{T}}(z) \sim_k \frac{9\sqrt{3}}{\sqrt{2\pi}} k^{-1/2} \left(\frac{1}{2}\right)^k$.

Proof. From (2.5.2) and the explicit expressions for $G(x)$ and $G(x,z)$ from (2.4.3) and (2.4.4) we can directly derive an explicit expression for $R_{\mathcal{G}^*}(z)$.

$$R_{\mathcal{G}^*}(z) = \frac{180z - 144 - 21z^3 - 24z^2 + (5z^2 - 26z + 24)\sqrt{3\,(2+z)\,(6-5z)}}{(1-z)\left(3z^2 - 36z + 36 + (6-5z)\sqrt{3\,(2+z)\,(6-5z)}\right)}\,.$$

Since $\mathcal{G}^*$ is not closed under re-rooting, we cannot directly apply (2.9) to obtain $E_{\mathcal{G}^*}(z)$. Still, by linearity of expectation, we obtain with $\rho := \rho_{\mathcal{G}}^* = \rho_{\mathcal{G}} = \rho_{\mathcal{H}} = \frac{1}{12}$ from (2.5.1) that

$$E_{\mathcal{G}^*}(z) = \frac{G^*(\rho)H(\rho)\,E_{\mathcal{G}^*\times\mathcal{H}}(z) + H(\rho)\,E_{\{\bullet\!\circlearrowleft\}\circ\mathcal{H}}(z) + H(\rho)\,E_{\{\bullet\bullet\}\circ\mathcal{H}}(z)}{G(\rho)}\,. \tag{2.5.7}$$

For maps drawn in the Boltzmann model from $\mathcal{G}^* \times \mathcal{H}$, the expected number of vertices of a given degree is that of the map from $\mathcal{G}^*$ (without root vertices), plus that of the map from $\mathcal{H}$ (again without root vertices), plus the expected number of root vertices that are of the given degree. Thus,

$$E_{\mathcal{G}^*\times\mathcal{H}}(z) = \big(E_{\mathcal{G}^*}(z) - 2\,R_{\mathcal{G}^*}(z)\big) + \big(E_{\mathcal{H}}(z) - 2\,R_{\mathcal{H}}(z)\big) + 2\,R_{\mathcal{G}^*}(z)\,R_{\mathcal{H}}(z)\,. \tag{2.5.8}$$

Maps in $\{\bullet\!\circlearrowleft\}\circ\mathcal{H}$ are composed out of two maps drawn from $G+\{\bullet\}$ in the Boltzmann model that are attached to each other at their root vertices and separated by an additional root edge. Thus for maps drawn in the Boltzmann model from $\{\bullet\!\circlearrowleft\}\circ\mathcal{H}$, the expected number of vertices of a given size is twice the expected number of those vertices in a map from $\mathcal{G}+\{\bullet\}$ (without the root vertex) plus the probability that the new root vertex composed of the two old root vertices and the new root edge has the requested degree. Thus,

$$E_{\{\bullet\!\circlearrowleft\}\circ\mathcal{H}}(z) = 2\,E_{\mathcal{G}+\{\bullet\}}(z) - 2\,R_{\mathcal{G}+\{\bullet\}}(z) + z^2\,R_{\mathcal{G}+\{\bullet\}}(z)^2\,. \tag{2.5.9}$$

Since $\{\bullet\!\rightarrow\!\bullet\}\circ\mathcal{H} = \mathcal{H}$, we have $E_{\{\bullet\bullet\}\circ\mathcal{H}}(z) = E_{\mathcal{H}}(z)$. If we substitute this, (2.5.8), and (2.5.9) in (2.5.7) and solve for $E_{\mathcal{G}^*}(z)$, we obtain an expression in $R_{\mathcal{G}^*}(z)$, $E_{\mathcal{G}}(z)$, $R_{\mathcal{H}}(z)$, $E_{\mathcal{G}+\{\bullet\}}(z)$, and $R_{\mathcal{G}+\{\bullet\}}(z)$, that is, all functions are given explicitly. Putting everything together yields an explicit expression for $E_{\mathcal{H}^*}(z)$ which we omit.

To obtain the asymptotic form of the coefficients first note that the function $R_{\mathcal{G}^*}(z)$ is strictly increasing for $z \in [0, \frac{6}{5}]$, and that $R_{\mathcal{G}^*}(\frac{6}{5}) = 2$. Hence, $R_{\mathcal{G}^*}$ is uniquely invertible in $[0, \frac{6}{5}]$, the inverse being

$$F(z) = \frac{20 + 46z + 17z^2 - (2-z)\sqrt{(50-z)\,(2-z)}}{34 + 26z + 16z^2}\,.$$

We thus obtain

$$D_{\mathcal{T}^\circ}(z) = 9\,D_{\mathcal{G}^*}(F(z)) - E_{\mathcal{G}^*}(F(z)) + 2\,R_{\mathcal{G}^*}(F(z)).$$

Now, as $D_{\mathcal{G}^*} = D_{\mathcal{G}}, E_{\mathcal{G}^*}, R_{\mathcal{G}^*}$ and F are given explicitly, we obtain an explicit expression for $D_{\mathcal{B}}(z)$. We see that the dominant singularity of $D_{\mathcal{B}}(z)$ is at 2, and that

$$D_{\mathcal{T}^\circ}(z) = 9\sqrt{\tfrac{3}{2}}\left(1 - \tfrac{z}{2}\right)^{-1/2} + \mathcal{O}(1) \text{ as } z \to 2.$$

The proof finishes by applying the Transfer Theorem (Theorem 2.5) to the above local expansion. $\square$

2.6. Loopless and Bridgeless Maps

For random loopless maps and random bridgeless maps we derive similar concentration results as for biconnected maps and c-nets. Note that the planar dual of a loopless map is a bridgeless map and vice versa, hence we only need to investigate one of the two classes.

Theorem 2.21. *Let $\varepsilon > 0$, $k \in \mathbb{N}$. Let L_n be a map drawn uniformly at random from the class $\mathcal{L}_n$ of all loopless maps (or, equivalently, bridgeless maps) with n edges. Then, uniformly for sufficiently large n*

$$\Pr[\deg(k; \mathsf{L}_n) \in (1 \pm \varepsilon)\, \ell_k\, n] = 1 - o(1)\,, \tag{2.6.1}$$

where $\ell_k = [z^k]D_{\mathcal{L}}(z)$ and the explicit expression of $D_{\mathcal{L}}(z)$ can be derived from (2.6.3). *Furthermore,*

$$\ell_k \sim_k \tfrac{3}{4\sqrt{\pi}}\, k^{-1/2} \left(\tfrac{3}{4}\right)^k .$$

Let $\mathcal{L}$ be the class of loopless maps. We define the *loopless core* of a map $G \in \mathcal{G}$ by removing all *maximal* loops and their interior from G (where a loop is maximal if it not contained within any other loop). If the root of G is a maximal loop, then the loopless core is empty. Tutte (1963) showed the following combinatorial relation.

Lemma 2.22 (Tutte (1963)). *Let $\mathcal{D} \subset \mathcal{M}$ be the class of general maps that have an empty loopless core. The classes $\mathcal{M}$ and $\mathcal{L}$ satisfy the relation*

$$\mathcal{M} = \mathcal{L} \circ \mathcal{H} + \mathcal{D}\,,$$

where

$$\mathcal{H} = \textsc{Seq}(\mathcal{M} + \{\bullet\}) \times \{\bullet\!\!\longrightarrow\!\!\bullet\} \times \textsc{Seq}(\mathcal{M} + \{\bullet\})$$

and $\textsc{Seq}(\mathcal{X})$ denotes a possibly empty sequence of elements in $\mathcal{X}$.

In words, any map in $\mathcal{M}$ has either an empty loopless core, or it is obtained from a loopless map substituting every edge by a map in $\mathcal{H}$. Here, $\mathcal{H}$ consists of all maps that are composed by attaching an arbitrary number of general maps that are enclosed in loops to the two endpoints of a single (root) edge.

By construction, $\mathcal{H}$ is closed under inversion of the orientation of the root edge, and moreover, Table 2.1 guarantees that the generating functions $G(x)$ and $L(x)$ are singular with exponent $3/2$. Finally, a straightforward calculation shows that also $H(x,z) = xz\,(1-x(G(x)+1))^{-1}\,(1-xz^2(G(x,z)+1))^{-1}$ is singular with exponent $3/2$, admitting the expansion

$$H(x) = \frac{27}{256} - \frac{81}{512}\left(1 - \tfrac{x}{\rho_{\mathcal{G}}}\right) + \frac{27}{512}\left(1 - \tfrac{x}{\rho_{\mathcal{G}}}\right)^{3/2} + \mathcal{O}\big(\big(1 - \tfrac{x}{\rho_{\mathcal{G}}}\big)^2\big), \tag{2.6.2}$$

as $x \to \rho_{\mathcal{H}} = \rho_{\mathcal{G}} = \frac{1}{12}$.

The above facts together with Theorem 2.1 imply that the preconditions of Theorem 2.12 are satisfied. So, we obtain Theorem 2.2, where the function $D_{\mathcal{B}}(z)$ is given by the relation

$$\frac{3}{2}\, D_{\mathcal{G}}(z) = E_{\mathcal{H}}(z) - 2R_{\mathcal{H}}(z) + D_{\mathcal{B}}\big(R_{\mathcal{H}}(z)\big), \tag{2.6.3}$$

and $D_{\mathcal{G}}(z)$ is given in Theorem 2.1. Since we know $H(x,z)$ explicitly, the functions $R_{\mathcal{H}}(z)$ can be derived straightforwardly. The function $E_{\mathcal{H}}(z)$ can also be calculated explicitly by decomposing $\mathcal{H}$ to its components from $\mathcal{M}$. In particular, a simple calculation yields that

$$E_{\mathcal{H}}(z) = \tfrac{1}{4}\left(E_{\mathcal{M}+\{\bullet\}}(z) - 2R_{\mathcal{M}+\{\bullet\}}(z)\right) + 2R_{\mathcal{H}}(z)\,.$$

From this, we infer that the expansion of $D_{\mathcal{L}}(z)$ at its dominant singularity 4/3 is

$$D_{\mathcal{L}}(z) = \tfrac{4}{3}\left(1 - \tfrac{3z}{4}\right)^{-1/2} + \mathcal{O}(1) \text{ as } z \to 4/3\,.$$

We apply the Transfer Theorem (Theorem 2.5) to the above expansion and obtain Theorem 2.21.

Indication of source. The content of this chapter has been previously published in the *Proceedings of SODA '10* (Johannsen and Panagiotou (2010)).

3 Counting Defective Parking Functions

A car park consists of n numbered spaces in a line. The drivers of m cars have independently chosen their favorite parking spaces. Each driver arrives at the car park and proceeds to his chosen space, parking there if it is free. If the chosen space is occupied, the driver continues on towards the larger-numbered spaces and takes the first available space if any; if no such space is available, the driver leaves the car park and goes home. What is the probability that everybody parks successfully? Equivalently, how many of the n^m sequences of choices by the drivers lead to everyone parking? (Such a sequence is called a *parking function*.)

This problem was first raised in the 1960s in connection with hashing (Konheim and Weiss (1966)). In the case of $m = n$, a short and elegant proof of the formula $(n+1)^{n-1}$ was given by Pollak (Foata and Riordan (1974)). From these beginnings, a substantial theory of parking functions has been developed, with links to trees (as one would expect from the formula above) and priority queues (Gilbey and Kalikow (1999)), partitions (Stanley (1997)), and representation theory (Haiman (1994)). More recently, generalizations of parking functions have found application in areas like the modelling of percolation (Majumdar and Dean (2002)), the Abelian sandpile model (Postnikov and Shapiro (2004)) and branching processes (Dumitriu et al. (2003)).

In this chapter, we are concerned with the probability that k drivers fail to park successfully. We call the corresponding assignments a *defective parking function* of *defect* k. Suppose that m cars attempt to park in a linear car park with n spaces according to the above rules; let $\mathrm{cp}(n, m, k)$ be the number of choices which result in exactly k drivers failing to park.

The concept is related to that of x–parking functions introduced by Pitman and Stanley (2002): For a tuple of integers $x = (x_1, \ldots, x_n)$ with $n \in \mathbb{N}$, an x–parking function is a sequence $(a_1, \ldots, a_n)$ whose ordered permutation $(b_1, \ldots, b_n)$ satisfies $b_i \leq x_1 + \ldots + x_i$ for all $1 \leq i \leq n$.

These generalized parking functions have been extensively studied. Pitman and Stanley (2002) related their number to the volume polynomials of certain types of polytopes and to that of plane partitions. Later, Kung and Yan (2003a,b) investigated moments of sums of their numbers by using Gončarov polynomials.

Classical parking functions correspond to $(1,1,\ldots,1)$–parking functions while defective parking functions of defect k correspond to $(n-(m-k)+1,1,\ldots,1,0,\ldots,0)$–parking functions which in addition are not $(n-(m-k),1,\ldots,1,0,\ldots,0)$–parking functions. For tuples $x=(a,b,\ldots,b,c,0,\ldots,0)$, an explicit formula was derived in Pitman and Stanley (2002) using the connection to polytope volumes and was reproven by Yan (2001) with combinatorial means.

We establish a new recurrence relation for the number of defective parking functions, allowing us to formulate an equation defining the corresponding three-variable generating function. Applying the kernel method (Flajolet and Sedgewick (2009); Prodinger (2004)), we solve this equation explicitly, and then extract the coefficients. We reobtain the fact that the cumulative totals turn out to be partial sums in *Abel's binomial formula* (Abel (1826)) as shown in Pitman and Stanley (2002) and Yan (2001) within the context of x–parking functions. In fact, the parking function approach may be used to prove special cases of this identity.

We then investigate the asymptotical behavior of defective parking for prominent cases. Spencer and Yan (2001) have studied asymptotics of parking functions with a defect equal to the difference between the number of cars and the number of spaces (the case in which all parking spaces end up being taken). We extend these results to the case of arbitrary defects. In particular, we include the case in which the number of cars is less than the number of spaces.

First, we show that, for fixed k and ℓ, the limit of $\operatorname{cp}(n,n+\ell,k)/\operatorname{cp}(n,n+\ell,0)$ exists, and compute its value. For example, the limiting value of $\operatorname{cp}(n,n,1)/\operatorname{cp}(n,n,0)$ is $2\mathrm{e}-3$.

To survey the limiting shape of the distribution, we need appropriate scaling, which turns out to be by the square root of n. We show that, if $m=n+\lfloor y\sqrt{n}\rfloor$, then the limiting probability of at most $\lfloor x\sqrt{n}\rfloor$ drivers failing to park is

$$\lim_{n\to\infty}\frac{1}{n^m}\sum_{k=0}^{\lfloor x\sqrt{n}\rfloor}\operatorname{cp}(n,m,k)=\begin{cases}1-e^{-2x(x-y)} & \text{if } x>y,\\ 0 & \text{otherwise;}\end{cases}$$

a surprisingly simple result, given the complicated form of the exact formula.

For $y=0$ (that is, in the case of $m=n$), this limiting distribution is the *Rayleigh distribution* with parameter $1/2$. (This occurs as the distribution of the length of a random vector in the plane whose coordinates are independent normal variables with standard deviation $1/2$. We do not know of a direct connection of this with our problem.)

We also investigate the limiting probability that all parking spaces are occupied. Obviously, for m strictly smaller than n this probability is zero. We show that for $m=\lfloor\lambda n\rfloor$ with fixed $\lambda\in\mathbb{R}^+$ and $k=m-n$,

$$\lim_{n\to\infty}\frac{\operatorname{cp}(n,m,k)}{n^m}=\begin{cases}0 & \text{if } \lambda\le 1,\\ 1-e^{-\lambda}\sum_{i\ge1}\frac{(\lambda i/e^{\lambda})^{i-1}}{i!} & \text{if } \lambda>1.\end{cases}$$

An alternative interpretation of the above car park problem involves a variation on the coupon collector problem. In the original problem, there are n distinct items.

n	$k=0$	1	2	3	4	5	6	7	8	9
1	1									
2	3	1								
3	16	10	1							
4	125	107	23	1						
5	1296	1346	436	46	1					
6	16807	19917	8402	1442	87	1				
7	262144	341986	173860	41070	4320	162	1			
8	4782969	6713975	3924685	1166083	176843	12357	303	1		
9	100000000	148717762	96920092	34268902	6768184	710314	34660	574	1	
10	2357947691	3674435393	2612981360	1059688652	256059854	36046214	2743112	96620	1103	1

Table 3.1. The table shows the numbers $\mathrm{cp}(n, n, k)$ for $n = 1, \ldots, 10$ and for $k = 0, \ldots, 9$ which count all car parking assignments of n cars to n spaces, such that k cars are not parked.

If a collector acquires random items, she will have approximately n/e duplicates after collecting the first n items and will need to collect about $n \log n$ items before she has a complete set. But suppose the items are of strictly decreasing value and she has the option of trading duplicate items; each item may be traded for any other one of lower value. Ideally, she trades duplicates against the next most valuable item she does not yet possess. Then, we show that she will receive only about $\sqrt{n}$ duplicates among the first n items and a complete collection already with $n\, f(n)$ items for any function $f\colon \mathbb{N} \to \mathbb{N}$ for which $\lim_{n\to\infty} f(n) = \infty$ holds.

We conclude this introduction with Pollak's lovely proof, adapted to the general case: for m cars and n spaces (with $m \leq n$), the number of ways in which every driver parks successfully is $(n+1-m)(n+1)^{m-1}$. To see this, consider a circular car park with $n+1$ spaces, for which the same rules apply. Now everyone will park successfully and there will be $n+1-m$ empty spaces; such a choice will be a parking function (for the original problem), if and only if space number $n+1$ is empty. By symmetry, this will happen in a fraction $(n+1-m)/(n+1)$ of the total number $(n+1)^m$ of choices. We will see later that our argument reproduces this result as an essential step in the working-out of the kernel method.

3.1. A Functional Equation

Let $\mathrm{cp}(n, m, k)$ be the number of assignments of m drivers to a car park with n spaces, that result in exactly k drivers leaving in the end, where the parking strategy of the drivers is as described in the introduction above. There are n^m such assignments. Then $\mathrm{cp}(n, m, k)$ can be concisely expressed as the number of functions $f\colon \{1, \ldots, m\} \to \{1, \ldots, n\}$ for which the set $f^{-1}(\{n+1-i, \ldots, n\})$ has a size of at most $k+i$ for all $i \in \{1, \ldots, n\}$, and at least one of these sets has a size of exactly $k+i$.

Values of $\mathrm{cp}(n, n, k)$, the case in which the number of drivers and the number of spaces coincide, can be found in Table 3.1.

We will now derive a recursion formula by transforming the parameters so that they are more suitable for our purpose. Let r be the number of spaces that end up unoccupied, and let s be the number of occupied spaces.

Definition 3.1. *For $r, s, k \in \mathbb{N}_0$ let $a(r, s, k)$ denote the number of choices for which r spaces remain unoccupied, s spaces are occupied in the end, and k people drive home.*

This obviously means that there are $n = r + s$ spaces in total, and that $m = k + s$ drivers arrive. Observe that $\mathrm{cp}(n, m, k) = a(n - m + k, m - k, k)$ is the number of assignments for the car parking problem with n parking spaces, m visitors, and k drivers going home. Correspondingly, $\mathrm{cp}(n, n, k) = a(k, n - k, k)$. Thus, finding a solution for $a(r, s, k)$ will yield a solution for the original problem. We extend this definition to all integers by setting $a(r, s, k) = 0$ whenever r, s, or k is smaller than 0. For the newly introduced numbers, we get the following recursive formula.

Lemma 3.2. *For $r, s, k \in \mathbb{N}_0$, the number of assignments of $s + k$ drivers to $r + s$ spaces, such that r spaces remain empty, s spaces are occupied, and k drivers leave, is recursively defined by*

$$a(r, s, k) = \begin{cases} 1 & \text{if } r = s = k = 0, \\ a(r-1, s, 0) + \sum_{i=0}^{k+1} \binom{s+k}{k+1-i} a(r, s-1, i) & \text{if } k = 0 \wedge (r > 0 \vee s > 0), \\ \sum_{i=0}^{k+1} \binom{s+k}{k+1-i} a(r, s-1, i) & \text{if } k > 0. \end{cases}$$

Proof. If $r = s = k = 0$, there exists exactly one assignment. Next, let $k > 0$, as in the third case. Since $k > 0$ drivers leave in the end, there are at least $k + 1$ drivers that arrive at parking space number $r + s$, counting the ones that actually chose it, as well as the ones that did not. An assignment of the $s + k$ drivers to the $r + s$ parking spaces satisfies this condition, if and only if for some $i \in \{1, \ldots, k + 1\}$ there are $k + 1 - i$ drivers that actually choose the last space and i drivers that arrive at the space, although they have not chosen it. For different values of i, the corresponding assignments of the drivers must differ.

There are $\binom{r+k}{k+1-i}$ ways to choose the $k + 1 - i$ drivers that actually pick the last space and $a(r, s-1, i)$ assignments of the remaining $s-1+i$ drivers to the first $r+s-1$ spaces, such that exactly i of them will arrive at the last space. Since these assignments are independent of the choice of $k+1-i$ drivers that choose the last space, the claimed recursion holds.

Finally, let $k = 0$ but $r > 0$ or $s > 0$. If some driver arrives at the last parking space, whether actually choosing it or just taking it due to lack of other spaces, the same recursion as for $k > 0$ holds. Otherwise, the last space will be empty and the number of assignments in which this happens is equal to the number of ways all s drivers can be assigned to the first $r + s - 1$ spaces, such that no driver has to leave. For this, there are exactly $a(r - 1, s, 0)$ ways. □

We restate the recursion formula from the previous lemma as

$$a(r, s, k) = \mathbf{1}_{\{r=s=k=0\}}(r, s, k) + \mathbf{1}_{\{k=0\}}(k)\, a(r-1, s, 0) + \sum_{i=0}^{k+1} \binom{s+k}{k+1-i} a(r, s-1, i)\,,$$

where $\mathbf{1}_{\{k=0\}}(k)$ and $\mathbf{1}_{\{r=s=k=0\}}(r, s, k)$ are the characteristic functions which are one, if $k = 0$ respectively $r = s = k = 0$ and zero, otherwise. Dividing both sides by $\binom{s+k}{k}$

yields

$$\frac{a(r,s,k)}{\binom{s+k}{k}} = \mathbf{1}_{\{r=s=k=0\}}(r,s,k)+\mathbf{1}_{\{k=0\}}(k)\,\frac{a(r-1,s,0)}{\binom{s+0}{0}}+\frac{s}{k+1}\sum_{i=0}^{k+1}\frac{\binom{k+1}{i}\,a(r,s-1,i)}{\binom{s-1+i}{i}}\,. \tag{3.1.1}$$

The previous equation suggests to represent $a(r, s, k)$ by a generating function which is ordinary in u and exponential in a combination of v and t.

Lemma 3.3. *Let A be the formal power series in the three variables u, w, and t defined by*

$$A(u,v,t) := \sum_{r,s,k\geq 0} a(r,s,k)\,u^r\,\tfrac{v^s t^k}{(s+k)!}\,.$$

Then A is the unique solution of

$$0 = (\tfrac{v}{t}e^t - 1)\,A(u,v,t) + (u - \tfrac{v}{t})\,A(u,v,0) + 1$$

in the ring of formal power series in u, v and t.

Proof. Multiplying both sides of (3.1.1) by $u^r\,\frac{v^s}{s!}\,\frac{t^k}{k!}$ and summing both sides over the parameters r, s, and k, followed by the usual manipulation, such as index-shifts and factorizing the product of $A(u, v, t)$ and e^t, shows that the definition of A and the equation stated in this lemma are equivalent. □

3.2. An Explicit Formula

We will now proceed to find an explicit formula for the coefficients of the generating function. In general, equations like the equation in Lemma 3.3 cannot be solved directly, since both $A(u, v, t)$ and $A(u, v, 0)$ are unknown. Instead, we resort to using the so-called kernel method (Flajolet and Sedgewick (2009); Prodinger (2004)). Writing the equation in Lemma 3.3 as

$$K(v,t)\,A(u,v,t) = (u - \tfrac{v}{t})\,A(u,v,0) + 1$$

with the *kernel* $K(v,t) = 1 - \frac{v}{t}e^t$, we solve for $A(u, v, 0)$ by setting the kernel equal to zero, which is here equivalent to finding a formal power series $t(v)$ for which $K(v, t(v)) = 0$. The solution to $t = ve^t$ is the well-known tree function $t = T(v)$ enumerating rooted trees on i labelled nodes, which is standardly expressed in terms of the Lambert W-function (Corless, Gonnet, Hare, Jeffrey, and Knuth (1996)) as $T(v) = -W(-v)$ and has series expansion

$$T(v) = \sum_{i=1}^{\infty}\frac{i^{i-1}}{i!}\,v^i\,.$$

Therefore

$$A(u,v,0) = \frac{e^{T(v)}}{1-ue^{T(v)}}\,, \tag{3.2.1}$$

which can be substituted into the equation of Lemma 3.3 to derive an explicit expression for $A(u, v, t)$. This proves the following lemma.

Lemma 3.4. *The generating function for the car parking problem is given by*

$$A(u,v,t)=\frac{1}{1-\frac{v}{t}e^t}+\frac{u-\frac{v}{t}}{1-\frac{v}{t}e^t}\,\frac{e^{T(v)}}{1-ue^{T(v)}}\,.$$

Applying Lagrange inversion to (3.2.1), we obtain the explicit expression

$$A(u,v,0)=\sum_{r,s\geq 0}(r+1)\,(r+s+1)^{s-1}\,u^r\,\tfrac{v^s}{s!}\,. \tag{3.2.2}$$

The coefficients of $A(u,v,0)$ have already been obtained in the introduction by a direct combinatorial method. It becomes apparent that we can express the car parking numbers $\mathrm{cp}(n,m,k)=a(n-m+k,m-k,k)$ in terms of the following sums, which also have a direct combinatorial interpretation.

Definition 3.5. *For $n,m,k\in\mathbb{N}_0$, let*

$$S(n,m,k):=\begin{cases} n^m & \text{if } k\leq m-n,\\ \displaystyle\sum_{i=0}^{m-k}\tbinom{m}{i}\,(n{-}m{+}k)\,(n{-}m{+}k{+}i)^{i-1}\,(m{-}k{-}i)^{m-i} & \text{otherwise.}\end{cases}$$

The car parking numbers then calculate as follows.

Theorem 3.6. *Let $n,m,k\in\mathbb{N}_0$. Then, the sum $S(n,m,k)$ counts the number of car parking assignments of m cars on n spaces, such that at least k cars do not find a parking space, that is,*

$$S(n,m,k)=\sum_{j=k}^{m}\mathrm{cp}(n,m,j)\,.$$

Equivalently, the car parking numbers $\mathrm{cp}(n,m,k)$ are given by

$$\mathrm{cp}(n,m,k)=S(n,m,k)-S(n,m,k+1)\,.$$

Proof. Expanding the explicit form of $A(u,v,t)$ in Lemma 3.4 with the help of (3.2.2) leads, after some lengthy manipulations, to

$$\begin{aligned} A(u,v,t)=&\sum_{s\geq 0}\sum_{k\geq 0}\tfrac{s^{s+k}\,v^s\,t^k}{(s+k)!}+\sum_{r\geq 1}\sum_{s\geq 0}\sum_{k\geq 0}\sum_{i=0}^{s}\tbinom{s+k}{i}\,r\,(r+i)^{i-1}\,(s-i)^{s+k-i}\,u^r\,\tfrac{v^st^k}{(s+k)!}\\ &-\sum_{r\geq 0}\sum_{s\geq 1}\sum_{k\geq 0}\sum_{i=0}^{s-1}\tbinom{s+k}{i}\,(r+1)\,(r+1+i)^{i-1}\,(s-1-i)^{s+k-i}\,u^r\,\tfrac{v^st^k}{(s+k)!}\\ =&\sum_{r,s,k\geq 0}\big(S(r+s,s+k,k)-S(r+s,s+k,k+1)\big)\,u^r\,\tfrac{v^st^k}{(s+k)!}\,. \end{aligned}$$

From this, we read off directly that

$$a(r,s,k)=S(r+s,s+k,k)-S(r+s,s+k,k+1)\,.$$

The statement of the theorem follows from the relation between $\mathrm{cp}(n,m,k)$ and $a(r,s,k)$. $\square$

Note that we can rewrite $S(n,m,k)$ for $k > m-n$ as

$$S(n,m,k) = \sum_{i=0}^{m-k} \binom{m}{i} \operatorname{cp}(n-m+k+i-1, i, 0)\,(m-k-i)^{m-i}\,.$$

This observation (Francesco (2008)) leads to a direct combinatorial proof of Theorem 3.6.

Alternative proof. If at least k cars do not find a parking space, then there are at least $\ell = n+k-m$ empty parking spaces. Assuming the ℓ-th empty space occurs at position $\ell+i$, there are i cars successfully parked in the $\ell+i-1$ spaces to the left of it, which is counted by $\operatorname{cp}(\ell+i-1, i, 0)$. Selecting these i cars out of all m cars can be done in $\binom{m}{i}$ different ways. The remaining $m-i$ cars are assigned to the $n-\ell-i$ rightmost spaces in $(n-\ell-i)^{m-i}$ different ways. Summing over all possible values of i leads to

$$S(n,m,k) = \sum_{i=0}^{n-\ell} \binom{m}{i} \operatorname{cp}(\ell+i-1, i, 0)\,(n-\ell-i)^{m-i}\,.$$

□

Another way to derive this theorem is to reinterpret defective parking functions in terms of x–parking functions as described in the introduction and then apply the results from Pitman and Stanley (2002).

3.3. Abel's Binomial Identity

In the last section we saw that the number of assignments of m cars to n spaces, such that at least k drivers fail to park, is the sum $S(n,m,k)$. Interestingly, $S(n,m,k)$ turns out to be a partial Abel-type sum as it appears in Abel's Binomial identity (Abel (1826)).

Lemma 3.7 (Abel's Binomial Identity).

$$\sum_{i=0}^{m} \binom{m}{i} a\,(a+i)^{i-1}\,(b-i)^{m-i} = (a+b)^m \qquad \textit{for all } a,b \in \mathbb{R}, m \in \mathbb{N}_0.$$

In fact, our approach gives a proof of this identity for the case $a, b, m \in \mathbb{N}_0$ and $b = m$. (Put $k = 0$, $a = n-m$ and $b = m$ in the defining equation of $S(n,m,k)$ in Definition 3.5). We use this identity to find the following short expressions of $S(n,m,k)$ and $S(n,n,k)$.

$$S(n,m,k) = \begin{cases} n^m & \text{if } k \le m-n, \\ n^m - \sum_{i=0}^{k-1} \binom{m}{i} (-1)^i\,(n-m+k)\,(k-i)^i\,(n+k-i)^{m-1-i} & \text{otherwise,} \end{cases}$$

$$S(n,n,k) = n^n - \sum_{i=0}^{k-1} \binom{n}{i} (-1)^i\,k\,(k-i)^i\,(n+k-i)^{n-1-i}\,.$$

We now give $S(n,n,k)$ for values of k close to zero or n:

$$\begin{aligned}
S(n,n,0) &= n^n\,,\\
S(n,n,1) &= n^n - (n+1)^{n-1}\,,\\
S(n,n,2) &= n^n - 2\,(n+2)^{n-1} + 2\,n\,(n+1)^{n-2}\,,\\
S(n,n,3) &= n^n - \tfrac{3}{2}\,n\,(n-1)\,(n+1)^{n-3} + 6\,n\,(n+2)^{n-2} - 3\,(n+3)^{n-1}\,,\\
&\vdots\\
S(n,n,n-3) &= 3^n + n\,(n-3)\,2^{n-1} + \tfrac{1}{2}\,n\,(n-1)^2\,(n-3)\,,\\
S(n,n,n-2) &= 2^n + n\,(n-2)\,,\\
S(n,n,n-1) &= 1\,,\\
S(n,n,n) &= \begin{cases} 1 & \text{if } n=0\,,\\ 0 & \text{otherwise}\,. \end{cases}
\end{aligned}$$

It follows that for $m = n + \ell$ with $k > \ell$ and $\ell \in \mathbb{Z}$

$$\lim_{n\to\infty} n \left(\frac{S(n,m,k)}{n^m} - 1 \right) = -\,(k-\ell) \sum_{i=0}^{k-1} \frac{(-1)^i}{i!}\,(k-i)^i\,e^{k-i} =: \phi(\ell,k)\,.$$

(This limit is trivially zero if $k \le \ell$.) This implies that for $\ell \le 0$ the limit

$$\lim_{n\to\infty} \frac{\mathrm{cp}(n,n+\ell,k)}{\mathrm{cp}(n,n+\ell,0)} = \lim_{n\to\infty} \frac{\phi(\ell,k) - \phi(\ell,k+1)}{\phi(\ell,0) - \phi(\ell,1)}$$

is finite (for $\ell > 0$ the denominator is zero). For example, we find

$$\lim_{n\to\infty} \frac{\mathrm{cp}(n,n,1)}{\mathrm{cp}(n,n,0)} = 2e - 3 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{\mathrm{cp}(n,n,2)}{\mathrm{cp}(n,n,0)} = 3e^2 - 8e + 7/2\,.$$

We conclude that in a random instance, we do not expect the number of drivers having to leave to be bounded by a constant.

Abel's binomial identity also gives us a first bound on $S(n,m,k)$. We obtain it by bounding the binomial coefficient appearing within the sum: Since for $k > m - n$, we know that

$$S(n,m,k) = \sum_{i=0}^{m-k} h(n,m,k) \binom{m-k}{i} (n-m+k)\,(n-m+k+i)^{i-1}\,(m-k-i)^{m-k-i}$$

with

$$h(n,m,k) = \frac{(m-k-i)^k\,(m-k-i)!\,m!}{(m-i)!\,(m-k)!} \le \frac{m!}{(m-k)!}\,.$$

It follows that

$$S(n,m,k) \le \frac{m!}{(m-k)!}\,n^{m-k}\,.$$

We have seen how the formula for $S(n,m,k)$, derived by Abel's identity, allows us to obtain asymptotic results for fixed values of k. In the following section, we analyze the case in which k grows as a function of n. Some of the upcoming results may be elementary proven using the previous inequality.

3.4. Asymptotics

For the parking problem with m cars and n spaces, the probability that at least k drivers cannot park their cars is $S(n,m,k)/n^m$ with $S(n,m,k)$ as defined in Definition 3.5. The case $k=0$ corresponds to partially filled hash tables and has been analyzed at length in Chassaing and Louchard (2002), with the most interesting asymptotic behavior obtained for $n-m=O(\sqrt{n})$. Similarly, we find non-trivial behavior in the regime where both $n-m$ and k are of order $O(\sqrt{n})$.

Theorem 3.8. *Let $x \in \mathbb{R}^+$ and $y \in \mathbb{R}$. Then, the limiting probability that in a random assignment of $n+\lfloor y\sqrt{n}\rfloor$ drivers to n spaces at least $\lfloor x\sqrt{n}\rfloor$ drivers fail to park is*

$$\lim_{n\to\infty} \frac{S(n, n+\lfloor y\sqrt{n}\rfloor, \lfloor x\sqrt{n}\rfloor)}{n^{n+\lfloor y\sqrt{n}\rfloor}} = \begin{cases} e^{-2x(x-y)} & \text{if } x > y, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. Let $p(n,m,k) = S(n,m,k)/n^m$. First note that the case $x \le y$ corresponds to the case $m \ge n+k$, where $p(n,m,k)=1$. The case $x>y$ corresponds to the case $m < n+k$, where

$$p(n,m,k) = \sum_{i=0}^{m-k} p(n,m,k,i)$$

with $p(n,m,k,i)$ given by

$$p(n,m,k,i) = \frac{1}{n^m}\binom{m}{i}(m-k-i)^{m-i}\,(n-m+k+i)^{i-1}\,(n-m+k)\,.$$

A straightforward but slightly tedious calculation establishes that for $\alpha \in [0,1]$

$$\lim_{n\to\infty} n\,p(n, n+y\sqrt{n}, x\sqrt{n}, \alpha n) = \frac{x-y}{\sqrt{2\pi\alpha^3(1-\alpha)}}\exp\left(-\frac{(x-(1-\alpha)y)^2}{2\alpha(1-\alpha)}\right).$$

As we have uniform convergence to a bounded limiting function for $\alpha \in [0,1]$, it is permissible to approximate $p(n, n+\lfloor y\sqrt{n}\rfloor, \lfloor x\sqrt{n}\rfloor)$ for large n by an integral as follows:

$$\begin{aligned}
\lim_{n\to\infty} p(n, n+\lfloor y\sqrt{n}\rfloor, \lfloor x\sqrt{n}\rfloor) &= \lim_{n\to\infty} \sum_{i=0}^{n-\lfloor x\sqrt{n}\rfloor} p(n, n+\lfloor y\sqrt{n}\rfloor, \lfloor x\sqrt{n}\rfloor, i) \\
&= \lim_{n\to\infty} \int_0^{1-x/\sqrt{n}} n\,p(n, n+y\sqrt{n}, x\sqrt{n}, \alpha n)\,d\alpha \\
&= \int_0^1 \frac{x-y}{\sqrt{2\pi\alpha^3(1-\alpha)}}\exp\left(-\frac{(x-(1-\alpha)y)^2}{2\alpha(1-\alpha)}\right)d\alpha\,.
\end{aligned}$$

Under the substitution $\alpha = \frac{u(x-y)}{x+u(x-y)}$, this integral simplifies to

$$\frac{1}{\sqrt{2\pi}}\int_0^\infty \sqrt{\frac{x(x-y)}{u^3}}\exp\left(-x\,(x-y)\,\frac{(1+u)^2}{2u}\right)du = \exp(-2x(x-y))\,.$$

$\square$

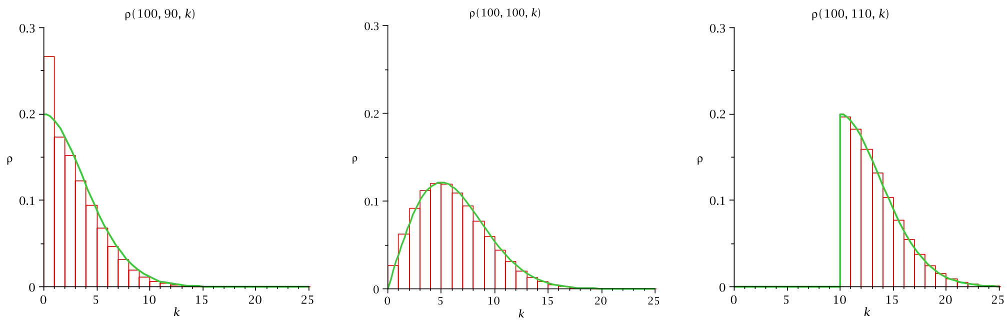

Figure 3.1. A comparison of the probabilities $\rho(n,m,k) = \mathrm{cp}(n,m,k)/n^m$, that m cars park randomly on n spaces such that k drivers fail to park, and their asymptotic approximation for $n = 100$ and $m = 90$ (left), $m = 100$ (middle) and $m = 110$ (right), respectively.

This theorem implies that for $m < n + k$, a good approximation is given by

$$\frac{\mathrm{cp}(n,m,k)}{n^m} \approx \frac{2}{n}\,(2k - m + n)\, e^{-2k(k-m+n)/n}\,.$$

Figure 3.1 shows a comparison between $\mathrm{cp}(n,m,k)/n^m$ and this approximation for the three qualitatively different scenarios $m < n$, $m = n$ and $m > n$.

Theorem 3.8 also shows that in the special case $m = n$, where the number of drivers and the number of parking spaces coincide, a random assignment will result in $\sqrt{n}$ drivers leaving. In particular we get the following corollary:

Corollary 3.9. *Let $k\colon \mathbb{N} \to \mathbb{N}$. Then, the limiting probability that in a random assignment of n drivers to n spaces at least $k(n)$ drivers fail to park is*

$$\lim_{n\to\infty} \frac{S(n,n,k(n))}{n^n} = \begin{cases} 0 & \text{if } \lim_{n\to\infty} k(n)/\sqrt{n} = \infty, \\ 1 & \text{if } \lim_{n\to\infty} k(n)/\sqrt{n} = 0. \end{cases}$$

Another question one may wish to ask is how the number m of cars needs to scale with the number n of parking spaces to fill the car park with a finite limiting probability, and when this probability reaches one. Similarly, from the viewpoint of the coupon collector, it is reasonable to ask how many coupons are needed to obtain a complete set.

Recall that the quantity $S(n,m,m-n+1)$ counts the number of car parking assignments of m cars on n spaces, such that at least $m - n + 1$ cars do not find a parking space, or, equivalently, such that at most $n - 1$ cars do find a parking space. Therefore, the probability that the car park is full is given by

$$\frac{\mathrm{cp}(n,m,m-n)}{n^m} = 1 - \frac{S(n,m,m-n+1)}{n^m}\,.$$

We find non-trivial behavior when m depends linearly on n as stated in the following theorem, a similar version of which can be found in Spencer and Yan (2001).

Theorem 3.10. *Let $\lambda \in \mathbb{R}^+$. Then, the limiting probability that in a random assignment of $\lfloor \lambda n \rfloor$ drivers to n spaces all spaces are occupied is*

$$\lim_{n\to\infty} \frac{\mathrm{cp}(n, \lfloor \lambda n \rfloor, \lfloor \lambda n \rfloor - n)}{n^{\lfloor \lambda n \rfloor}} = \begin{cases} 0 & \text{if } \lambda \leq 1, \\ 1 - \frac{1}{\lambda} T\left(\lambda e^{-\lambda}\right) & \text{if } \lambda > 1. \end{cases}$$

Proof. Let $p(n,m) = S(n,m,m-n+1)/n^m$. We have

$$p(n,m) = \frac{1}{n^m} \sum_{i=0}^{n-1} \binom{m}{i} (i+1)^{i-1} (n-1-i)^{m-i}.$$

Choosing $m = \lfloor \lambda n \rfloor$, we exchange summation with taking the limit and compute the limit term-wise, arriving at

$$\lim_{n\to\infty} p(n, \lfloor \lambda n \rfloor) = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} (i+1)^{i-1} e^{-\lambda(1+i)}. \tag{3.4.1}$$

Exchanging summation with taking the limit is justified by the Lebesgue dominated convergence theorem, as each term is bounded by the limiting expression, that is,

$$0 \leq \frac{1}{n^m} \binom{m}{i} (i+1)^{i-1} (n-1-i)_+^{m-i} \leq \frac{(m/n)^i}{i!} (i+1)^{i-1} e^{-(1+i)m/n}.$$

The sum in (3.4.1) converges for $|\lambda e^{-\lambda}| \leq 1$ (in particular for all positive λ). It can be expressed using the tree function $T(v)$ with $v = \lambda e^{-\lambda}$, and we find

$$\lim_{n\to\infty} p(n, \lfloor \lambda n \rfloor) = \frac{1}{\lambda} T\left(\lambda e^{-\lambda}\right).$$

Recalling that $t = T(v)$ solves $v = te^{-t}$, we find that determining $t = T\left(\lambda e^{-\lambda}\right)$ reduces to solving $te^{-t} = \lambda e^{-\lambda}$. For $\lambda \leq 1$, we find $t = \lambda$, however, for $\lambda > 1$ no such simplification is possible. □

Figure 3.2 shows a comparison between $\mathrm{cp}(n,m,m-n)/n^m$ with $m = \lfloor \lambda n \rfloor$ and finite values of n and the limiting curve given by $1 - T(\lambda e^{-\lambda})/\lambda$.

We conclude with a corollary of the previous theorem which emphasizes the threshold character of its statement. In this sense, the relation between the following corollary and the previous theorem reflects the relation between Corollary 3.9 and Theorem 3.8.

Since $\frac{1}{e^{\lambda-1}-\lambda}$ is an upper bound for $T(\lambda e^{-\lambda})/\lambda$ which converges to zero for $\lambda \to \infty$, almost all random assignments of $m = m(n)$ drivers to n parking spaces will result in all spaces being occupied if m/n tends to infinity.

Corollary 3.11. *Let $m\colon \mathbb{N} \to \mathbb{N}$. Then, the limiting probability that in a random assignment of $m(n)$ drivers to n spaces all spaces are occupied is*

$$\lim_{n\to\infty} \frac{\mathrm{cp}(n, m(n), m(n) - n)}{n^{m(n)}} = \begin{cases} 0 & \text{if } \lim_{n\to\infty} m(n)/n \leq 1, \\ 1 & \text{if } \lim_{n\to\infty} m(n)/n = \infty. \end{cases}$$

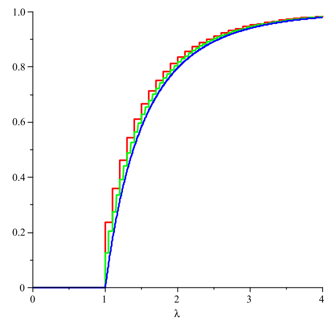

Figure 3.2. A comparison of the probabilities $\mathrm{cp}(n, m, m-n)/n^m$ for $n = 10$, $n = 20$, and the limiting curve given by $1 - T(\lambda e^{-\lambda})/\lambda$ from top to bottom, respectively. Here, $m = \lfloor \lambda n \rfloor$.

3.5. Conclusion

We have derived the generating function with coefficients $\mathrm{cp}(n, m, k)$ for the defective car parking numbers and have used it to solve the problem of their exact and asymptotic enumeration. They are closely connected to the x–parking functions for which numerous structural interpretations exist, and these interpretations therefore transfer. However, we expect applications of our results in different areas, since parking functions, with or without defect, naturally capture various time-dependent models.

For example, motivated by hashing with linear probing, a drop-push model for percolation was proposed in Majumdar and Dean (2002). Here, particles are dropped sequentially on a substrate, followed by the transport of the dropped particles via a pushing mechanism, caused by a local hard-core repulsion between particles on the substrate. If the transport is unidirectional, this is identical to the parking problem studied in this chapter.

Another example, the Abelian sandpile model, allows a decrease in the quantity of the system that corresponds to the number of cars. Such a decrease is not only possible but even necessary to prove certain stability properties of the system. Translated to parking functions, the notion of a defect naturally appears (Dhar (1990)).

A further potential field of application is queueing theory. For instance, the branching process described in Dumitriu et al. (2003) can by viewed in this context. With respect to queueing, parking spaces are interpreted as time slots, at each of which exactly one task (represented by one car) can be processed. Whereas in a branching process, at least one unprocessed task must exist at every point in time; there is no reason to forbid idle time in a more general queueing process.

Related results have been independently obtained by Panholzer (2008), notably a detailed description of limiting distributions.

Indication of source. The content of this chapter has been published in the *Electronic Journal of Combinatorics* (Cameron, Johannsen, Prellberg, and Schweitzer (2008))

Part II

Randomized Search Heuristics

4 Probabilistic Methods

In this chapter, we present existing as well as new probabilistic techniques as used in the analysis of random search heuristics. In this, we focus on the techniques applied in the next chapter and put them into the context of related probabilistic methods.

We start with a section that reviews well-known tail bounds like the Markov inequality, the Chernoff bounds, and the Azuma-Hoeffding inequality (which we already applied in Chapter 2).

The second section introduces drift analysis. The theorem on multiplicative drift (Theorem 4.5), a variant of which is published in Doerr, Johannsen, and Winzen (2010b), is the main method of analysis for the problems discussed in the next chapter.

The third section is concerned with the random process known as to as *gambler's ruin*. We discuss how the understanding of a generalized version of this process can be applied in the analysis of randomized search heuristics. The content of this section has been published in Happ, Johannsen, Klein, and Neumann (2008).

In the fourth and last section, we develop a theorem on dominance of stochastic processes. We will see that under proper assumptions we are can bound a sequence of dependent random variable by a sequence of independent random variables.

4.1. Preliminaries

In this section we state three well-known tail inequalities (see, for example, McDiarmid (1989) and Alon and Spencer (2008)) which we apply repeatedly in this thesis, for example, in Section 2.3 of Chapter 2 and in the next chapter.

The Markov inequality is a basic yet for many applications sufficient bound that links the tail probability of a random variable with its expectation.

Theorem 4.1 (Markov Inequality). *Let X be a non-negative random variable. Then, for all $t \in \mathbb{R}^+$,*

$$\Pr\left[X \geq t\right] \leq \frac{\mathrm{E}[X]}{t}\,.$$

If we know more about the properties of a random variable, we are able to give better bounds. The Chernoff bounds provide us with sharp concentration statements for sums of independent and identically-distributed binary random variables.

Theorem 4.2 (Chernoff Bounds). *Let $X_1, \ldots, X_n$ be independent and identically-distributed random variables over $\{0,1\}$. Furthermore, let $X := \sum_{i=1}^{n} X_i$ and let $\mu := \mathrm{E}[X]$. Then, for all $\delta \in [0,1]$,*

$$\Pr\left[X \leq (1-\delta)\mu\right] \leq e^{-\frac{\delta^2\mu}{2}} \text{ and}$$
$$\Pr\left[X \geq (1+\delta)\mu\right] \leq e^{-\frac{\delta^2\mu}{3}}\,.$$

The following version of the *Azuma-Hoeffding's inequality* (McDiarmid (1989)) provides us with sharp concentration for a random variable over a product probability space if the effect of the single coordinates is small.

Theorem 4.3 (Azuma-Hoeffding). *Let $X_1, \ldots, X_n$ be independent random variables with X_i taking values in a set A_i for each $i \in \{1, \ldots, n\}$. Suppose there exist a (measurable) function $f\colon \prod_{i=1} A_i \to \mathbb{R}$ and constants $c_1, \ldots, c_n \in \mathbb{R}$ such that for all $i \in \{1, \ldots, n\}$, the function f satisfies*

$$|f(x) - f(x')| \leq c_i$$

whenever the vectors x and x' differ only in the i-th coordinate. Let X be the random variable $X := f(X_1, \ldots, X_n)$ and $\mu = \mathrm{E}[X]$. Then, for all $t \in \mathbb{R}^+$,

$$\Pr\left[|X - \mathrm{E}[X]| \geq t\right] \leq 2e^{-2t^2/\sum_{i=1}^{n} c_i^2}\,.$$

4.2. Drift Analysis

The *drift* of a sequence $\{X^{(t)}\}_{t\in\mathbb{N}}$ of random variables over $\mathbb{R}_0^+$ is its expected change over time. More precisely, the drift of $\{X^{(t)}\}_{t\in\mathbb{N}}$ at time $t \in \mathbb{N}$ is the expected value of $\Delta^{(t)} := X^{(t)} - X^{(t+1)}$.

For such a sequence and a value $x_0 \in \mathbb{R}_0^+$, the *first hitting time with respect to x_0* is the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $X^{(t)} = x_0$. In our case, x_0 is typically zero. Drift analysis aims at deriving knowledge on the first hitting time of a sequence from knowledge on its drift.

In this section, we give an overview over probabilistic techniques for the study of randomized search heuristics that are based on drift analysis. Most of these techniques are based on the work of He and Yao (2001, 2004) and go back to the work of Hajek

(1982) and have been adapted to the specific setting we encounter when analysing randomized search heuristics.

Several of the results we present in this section have been formulated for sequences of random variables that describe Markov chains (He and Yao (2004); Giel and Lehre (2006); Happ et al. (2008); Oliveto and Witt (2010)). However, this limitation is not necessary. All proofs in these works are based on He and Yao (2001) and on Hajek (1982), neither of both using any Markov chain arguments (nor do the proofs of the derived results). Instead, the proofs (implicitly) use the fact that the preconditions of the respective statements guarantee that the considered sequences are submartingales or supermartingales (Hajek (1982)). We therefore can formulate all results in this section for general sequences.

4.2.1. Upper Bounds on the Expected First Hitting Time

The first theorem of this subsection (He and Yao (2001)) enables us to give upper bounds on the first hitting times of random processes for which we know constant lower bounds on the drift.

Theorem 4.4 (Constant drift). *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite state space $\mathcal{S} \subseteq \mathbb{R}_0^+$. Furthermore, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $X^{(t)} = 0$. Suppose that there exists a constant $C > 0$ such that*

$$\mathrm{E}\big[X^{(t)} - X^{(t+1)} \,\big|\, t < T\big] \geq C\,.$$

Then,

$$\mathrm{E}\big[T \,\big|\, X^{(0)}\big] \leq \frac{X^{(0)}}{C}\,.$$

In the random processes we will consider in the next chapter, the drift is usually bounded by a multiple of $X^{(t)}$ instead of a mere constant. In the corresponding upper bounds on the expected first hitting times (Theorem 5.4, Theorem 5.10, Theorem 5.26, and Theorem 5.36), This property accounts for the logarithmic factor. A variant of the following theorem is published in Doerr, Johannsen, and Winzen (2010b).

Theorem 4.5 (Multiplicative drift). *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite state space $\mathcal{S} \subseteq \mathbb{R}_0^+$ and let $x_{\min} := \min\{x \in \mathcal{S} \colon x > 0\}$. Furthermore, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $X^{(t)} = 0$.*

Suppose that there exists a constant $C > 0$ such that

$$\mathrm{E}\big[X^{(t)} - X^{(t+1)} \,\big|\, X^{(t)}\big] \geq C\, X^{(t)}$$

holds for all $t < T$. Then, if $X^{(0)} \neq 0$,

$$\mathrm{E}\big[T \,\big|\, X^{(0)}\big] \leq \frac{1}{C}\left(1 + \ln \frac{X^{(0)}}{x_{\min}}\right)\,.$$

The previous theorem is a corollary of the following, more general result, which is derived from Theorem 4.4 and builds on the results in He and Yao (2004).

Theorem 4.6 (Variable drift). *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite state space $\mathcal{S} \subseteq \mathbb{R}_0^+$ and let $x_{\min} := \min\{x \in \mathcal{S} : x > 0\}$. Furthermore, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $X^{(t)} = 0$. Suppose that there exists a continuous and monotone increasing function $h\colon \mathbb{R}_0^+ \to \mathbb{R}^+$ such that*

$$\mathrm{E}\big[X^{(t)} - X^{(t+1)} \,\big|\, X^{(t)}\big] \geq h(X^{(t)})$$

holds for all $t < T$. Then,

$$\mathrm{E}\big[T \,\big|\, X^{(0)}\big] \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X^{(0)}} \frac{1}{h(x)}\,\mathrm{d}x\,.$$

Proof. Let $g\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ be the function defined by

$$g(z) := \begin{cases} \frac{z}{h(x_{\min})} & \text{if } 0 \leq z < x_{\min}\,, \\ \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{z} \frac{1}{h(x)}\,\mathrm{d}x & \text{if } z \geq x_{\min}\,. \end{cases}$$

The function g is strictly monotone increasing, continuous on all $z \in \mathbb{R}^+$, and right-continuous at $z = 0$. Moreover, g is differentiable on $\mathbb{R}^+$ with

$$g'(z) = \begin{cases} \frac{1}{h(x_{\min})} & \text{if } 0 \leq z < x_{\min}\,, \\ \frac{1}{h(z)} & \text{if } z \geq x_{\min}\,. \end{cases}$$

We now apply the mean-value theorem and show that for all $x, y \in \mathbb{R}_0^+$ with $x \geq x_{\min}$ it holds that

$$g(x) - g(y) \geq \frac{x - y}{h(x)}\,. \tag{4.2.1}$$

If $x = y$ equation (4.2.1) holds trivially.
If $x < y$ then by the mean-value theorem there exists a $\xi \in (x, y)$ such that

$$g'(\xi) = \frac{g(y) - g(x)}{y - x}\,.$$

If $x > y$ then by the mean-value theorem there exists a $\xi \in (x, y)$ such that

$$g'(\xi) = \frac{g(x) - g(y)}{x - y}\,.$$

In both cases, after manipulation of the inequalities, equation (4.2.1) follows because g' is monotone decreasing.

Since g is strictly monotone increasing, its inverse g^{-1} exists. Moreover, the variable T from the theorem describes also the first point in time $t \in \mathbb{N}$ for which $g\big(X^{(t)}\big)$ is zero. Thus, for all $t < T$ it holds that

$$\mathrm{E}\big[g(X^{(t)}) - g(X^{(t+1)}) \,\big|\, g(X^{(t)})\big] \geq \mathrm{E}\Big[\tfrac{X^{(t)} - X^{(t+1)}}{h(X^{(t)})} \,\Big|\, X^{(t)}\Big] \geq 1\,.$$

The proof concludes with application of the drift theorem for upper bounds (Theorem 4.4). We obtain

$$\mathrm{E}\big[T \,\big|\, X^{(0)}\big] = \mathrm{E}\big[T \,\big|\, g(X^{(0)})\big] \leq g(X^{(0)})$$

and the result follows by substituting $g(X^{(0)})$. □

4.2.2. Lower Bounds on the Expected First Hitting Time

In the previous subsection, we have seen how a positive lower bound on the drift implies upper bounds on the expected first hitting time. In this section, we see how lower bounds on the expected first hitting time can be derived from upper bounds on the drift.

First, we state an analogue of Theorem 4.4 for lower bounds on first hitting times.

Theorem 4.7 (Constant drift (lower bound)). *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite state space $\mathcal{S} \subseteq \mathbb{R}_0^+$. Furthermore, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $X^{(t)} = 0$. Suppose that there exists a constant $C > 0$ such that*

$$\mathrm{E}\big[X^{(t)} - X^{(t+1)} \,\big|\, t < T\big] \leq C\,.$$

Then,

$$\mathrm{E}\big[T \,\big|\, X^{(0)}\big] \geq \frac{X^{(0)}}{C}\,.$$

This theorem deals with the situation where the drift is small but potentially positive. If the drift is guaranteed to be negative, we can derive *exponential* lower bounds on the first hitting time that hold with high probability. The following theorem (Giel and Lehre (2006)) is based on the work of Hajek (1982) and He and Yao (2001).

Theorem 4.8. *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a state space $\mathcal{S}$ and $g\colon S \to \mathbb{R}_0^+$ a function mapping each state to a non-negative real number. Pick two real numbers $a(n)$ and $b(n)$ depending on a parameter n such that $0 \leq a(n) < b(n)$ holds. Let the random variable T denote the first point in time $t \geq 0$ satisfying $g(X^{(t)}) \leq a(n)$. If there are constants $\lambda > 0$ and $D \geq 1$ and a polynomial $p(n) > 0$ such that the four conditions*

1. *$g(X^{(0)}) \geq b(n)$,*
2. *$b(n) - a(n) = \Omega(n)$,*
3. *$\mathrm{E}\big[e^{-\lambda\,(g(X^{(t+1)}) - g(X^{(t)}))} \,\big|\, X^{(t)}\big] \leq 1 - \frac{1}{p(n)}$ for all $X^{(t)}$ with $a(n) < g(X^{(t)}) < b(n)$ and $t \geq 0$,*
4. *$\mathrm{E}\big[e^{-\lambda\,(g(X^{(t+1)}) - b(n))} \,\big|\, X^{(t)}\big] \leq D$ for all $t \geq 0$ and $X^{(t)}$ with $g(X^{(t)}) \geq b(n)$*

hold then for all time bounds $B \geq 0$

$$P\left[T \leq B\right] \leq e^{\lambda\,(a(n) - b(n))}\, B\, D\, p(n)$$

holds.

Whenever we apply the previous result directly, we have to analyze the moment generating function $\mathrm{E}\left[e^{\lambda\,(g(X^{(t)})-g(X^{(t+1)}))} \,\middle|\, X^{(t)}\right]$ of $g(X^{(t)}) - g(X^{(t+1)})$ conditioned on $X^{(t)}$. This is avoided in the following special case (Oliveto and Witt (2010)) of the previous theorem. It is applied in the proof of Theorem 5.19 in the next chapter to prove an exponential lower bound on the runtime of the (1+1) Evolutionary Algorithm on the maximum matching problem.

Theorem 4.9 (Negative drift). *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite state space $\mathcal{S}$. Let $g\colon \mathcal{S} \to \{0,\dots,n\}$ with $n \in \mathbb{N}$ be a potential function on $\mathcal{S}$. Moreover, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $g(X^{(t)}) = 0$.*

Suppose that there exists three constants $\delta > 0$, $\varepsilon > 0$, $r > 0$ such that for all $t \geq 0$

1. $\mathrm{E}\left[g(X^{(t)}) - g(X^{(t+1)}) \,\middle|\, g(X^{(t)}) = k\right] \leq -\varepsilon$ *for all* $0 < k < n$,
2. $\Pr\left[g(X^{(t)}) - g(X^{(t+1)}) = \ell \,\middle|\, g(X^{(t)}) = k\right] \leq \left(\frac{1}{1+\delta}\right)^{\ell - r}$ *for all* $0 < k \leq n$, $\ell \geq 1$.

Then there exist constants $\alpha, \beta > 0$ such that

$$\Pr\left[T \leq \mathrm{e}^{\alpha\, n} \,\middle|\, g(X^{(0)}) = n\right] \leq \mathrm{e}^{-\beta\, n}\,.$$

4.3. The Gambler's Ruin

The *gambler's ruin*, a stochastic process, is of special interest to us since it can be used to describe the behavior of a randomized search heuristics in certain situations (see, for example, Happ et al. (2008)).

In the classical gambler's ruin problem (see, for example, Feller (1968)), the following process is studied. Let x be the amount of dollars a gambler owns at the beginning of a series of bets. With every bet the gambler wins one dollar with probability p and looses one dollar with probability $q = 1 - p$. In this setting we are interested in the probability that the gambler wins $z - x$ dollars in total, that is, that his capital reaches an amount of $z > x$ dollars before it attains the amount of zero dollars.

Theorem 4.10 (Gambler's Ruin). *Let p be the probability of winning one dollar and $q = 1 - p$ be the probability of loosing one dollar in a single bet and let $\delta = q/p$. Starting with x dollars, the probability of reaching $z > x$ dollars before attaining zero dollars is*

$$p_x = \begin{cases} \frac{x}{z} & \text{if } p = q = 1/2\,, \\ \frac{\delta^x - 1}{\delta^z - 1} & \text{otherwise.} \end{cases}$$

The previous theorem (see, for example, Feller (1968)) on the gambler's ruin process is well suited to prove lower bounds on the runtime behavior of the search heuristic *Randomized Local Search* which we have described in Section 1.2 of the introduction.

Consider a search space with a neighborhood structure represented by a graph. Let $X^{(t)}$ be the distance between the currently best candidate solution of Randomized Local Search and a specific search point (vertex in the graph). In each iteration, Randomized Local Search takes (at most) one step in the graph. Therefore, $X^{(t)}$ changes by at most one and we can reduce the analysis of this change to the analysis of the gambler's ruin process.

Theorem 4.11. *Let $n \in \mathbb{N}$ and $\{X^{(t)}\}_{t \in \mathbb{N}}$ be a sequence of random variables over the state space $\mathbb{N}_0$ with $X^{(t+1)} - X^{(t)} \in \{-1, 0, 1\}$ for all $t \geq 0$. Let $a, b \in \mathbb{R}$ be constants with $0 \leq a < b \leq 1$ and let the random variable T denote the first point in time $t \in \mathbb{N}$ that satisfies $X^{(t)} \leq an$.*

If there exists a constant $\delta > 1$ such that the two conditions

(a) $\Pr[X^{(0)} \geq bn] = 1 - 2^{-\Omega(n)}$ *and*

(b) $\Pr[X^{(t+1)} - X^{(t)} = 1 \mid X^{(t)}] \geq \delta \Pr[X^{(t+1)} - X^{(t)} = -1 \mid X^{(t)}]$ *holds for all $t \geq 0$ and $an < X^{(t)} < bn$,*

hold then $T \geq \delta^{1/3\,(b-a)\,n}$ with probability at least $1 - 2^{-\Omega(n)}$.

Proof. For $t \geq 0$ let $X^{(t)}$, $0 \leq a < b \leq 1$, and $\delta > 1$ be defined as above and suppose that conditions (a) and (b) hold.

Condition (a) guarantees that if we condition on the event that $X^{(0)} \geq bn$ then the theorem still holds with probability $1 - 2^{-\Omega(n)}$. Thus, we may assume that $X^{(0)} \geq bn$.

Let $p = 1/(1+\delta)$ and $q = \delta/(1+\delta)$. By condition (b), we have

$$\begin{aligned} \Pr[X^{(t+1)} - X^{(t)} &= -1 \mid X^{(t)}] \leq p \quad \text{and} \\ \Pr[X^{(t+1)} - X^{(t)} &= 1 \mid X^{(t)}] \geq q \end{aligned}$$

for all $t \geq 0$ and $X^{(t)}$ such that $an < X^{(t)} < bn$.

Let $t \in \mathbb{N}$ with $t < T$. By Corollary 4.14, there exists a independent random variables $\Delta^{(0)}, \ldots, \Delta^{(t)}$ such that $\Pr[\Delta^{(t)} = -1] = p$ and $\Delta^{(t)} \leq X^{(t+1)} - X^{(t)}$ holds. Then, $\widetilde{X}^{(t)} := \sum_{s=0}^{t} \Delta^{(t)}$ dominates $X^{(t)}$.

Since the previous statement holds for all $t \in \mathbb{N}$ with $t < T$, we can apply the Gambler's Ruin Theorem (Theorem 4.10) to $\{\widetilde{X}^{(t)}\}_{t \in \mathbb{N}}$ with values $z = \lfloor (b-a)\,n \rfloor$ and $x = \lfloor z/2 \rfloor$ and choose n sufficiently large such that $z - x \geq 4/9\,(b-a)\,n$. Since $\delta = q/p > 1$, the probability p_x of reaching z before attaining zero is at most $\delta^{-(z-x)}$.

Thus, starting with $X^{(0)} \geq bn$, the probability that $X^{(t)}$ reaches a value of at most an after passing the value $\lfloor bn \rfloor - x$ is at most $\delta^{-4/9\,(b-a)\,n}$. Given less then $\delta^{1/3\,(b-a)\,n}$ such tries, the probability to succeed is still at most $\delta^{-1/9\,(b-a)\,n}$. Hence, $T \geq \delta^{1/3\,(b-a)\,n}$ with probability $1 - 2^{-\Omega(n)}$. $\square$

Unlike Randomized Local Search, the (1+1) Evolutionary Algorithm (which we analyze in the next section) can perform more than one step on the graph that models the search space in each iteration (a description of the (1+1) Evolutionary Algorithm can be found in Section 1.2 of the introduction and a precise definition in the next chapter). For this situation, we provide a generalized version of the previous theorem, which is again a specialization of Theorem 4.8.

We may interpret the following result by a generalized gambler's ruin process. In this case, the gambler can win or loose $j \geq 1$ dollars with a certain probability in every step. Similar to the previous theorem, we derive a first hitting time result from the ratio between the probabilities of winning and losing j dollars for any fixed choice of j.

Theorem 4.12 (Generalized Gambler's Ruin). *Let $n \in \mathbb{N}$ and $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over the state space $\mathbb{N}_0$. For constants $a, b \in \mathbb{R}$ with $0 \le a < b \le 1$ let the random variable T denote the first point in time $t \in \mathbb{N}$ that satisfies $X^{(t)} \le an$.*

If there exist constants $\delta > 1$ and $C > 0$ such that the three conditions

(a) $\Pr[X^{(0)} \ge bn] = 1 - 2^{-\Omega(n)}$,

(b) $\Pr[X^{(t+1)} - X^{(t)} = j \mid X^{(t)}] \ge \delta^j \Pr[X^{(t+1)} - X^{(t)} = -j \mid X^{(t)}]$ *holds for all* $j \ge 1$, $t \ge 0$, *and* $an < X^{(t)} < bn$, *and*

(c) $\sum_{j\ge 1} \delta^j \Pr[X^{(t+1)} - X^{(t)} = -j \mid X^{(t)}] \le C$ *holds for all* $t \ge 0$ *and* $X^{(t)} \ge bn$

hold then $T \ge \delta^{1/3\,(b-a)\,n}$ with probability at least $1 - 2^{-\Omega(n)}$.

Proof. Suppose that conditions $(a)-(c)$ hold, and let $\Delta_t = X^{(t+1)} - X^{(t)}$.

We apply the Drift Theorem with the same random variables, $g = \mathrm{id}$, $a(n) = an$, $b(n) = bn$, $\lambda = \ln(\delta)/2$, $D = C + 1$, and $p(n) = (1 + \delta^{-1})/(1 - \delta^{-1/2})^2$. That is, $S = \mathbb{N}_0^+$, $g(X) = X$, $e^{\lambda} = \delta^{1/2} > 1$, and $p(n)$ is a strictly positive constant. Let $\mu_t := \mathrm{E}(e^{-\lambda\,(g(X^{(t+1)}) - g(X^{(t)}))} \mid X^{(t)})$, then

$$\mu_t = \sum_{j\in\mathbb{Z}} \delta^{-j/2} \Pr[\Delta_t = j \mid X^{(t)}]\,.$$

We check conditions 1.–4. of the Drift Theorem.

1. If the statement follows for $X^{(0)} \ge bn$, then it also follows if $X^{(0)} \ge bn$ with probability $1 - 2^{-\Omega(n)}$. Hence, condition 1 is satisfied by condition (a).
2. Clearly, $b(n) - a(n) = (b - a)n = \Omega(n)$.
3. Suppose that $t \ge 0$ and $X^{(t)}$ are such that $a(n) < g(X^{(t)}) < b(n)$, that is, $an < X^{(t)} < bn$. We replace $\Pr[\Delta_t = j \mid X^{(t)}]$ for all $j \in \mathbb{Z}$ such that μ_t increases. This is done in three steps:

   First, we suppose that $\Pr[\Delta_t = 0 \mid X^{(t)}] < 1$ for all $an < X^{(t)} < bn$. We can do so, because if $X^{(t)}$ takes a value such that $\Pr[\Delta_t = 0 \mid X^{(t)}] = 1$ then $\Pr[T \ge B] = 1$ for every $B \ge 0$ since the process never leaves this state.

   Second, we ignore whenever $X^{(t)} = X^{(t+1)}$ for some $t \ge 0$ in the definition of T. Clearly, this never increases T. Formally, we replace $\Pr[\Delta_t = 0 \mid X^{(t)}]$ by zero and $\Pr[\Delta_t = j \mid X^{(t)}]$ by $\Pr[\Delta_t = j \mid X^{(t)}]/(1 - \Pr[\Delta_t = 0 \mid X^{(t)}])$ for all $j \ne 0$. Thus,

   $$\mu_t \le \sum_{j\ne 0} \delta^{-j/2}\, \frac{\Pr[\Delta_t = j \mid X^{(t)}]}{1 - \Pr[\Delta_t = 0 \mid X^{(t)}]}\,.$$

   Third, since $\delta > 1$, the right hand-side of this inequality never decreases if we increase $\Pr[\Delta_t = -j \mid X^{(t)}]$ by some amount and decrease $\Pr[\Delta_t = 0 \mid X^{(t)}]$ by the same amount for any $j \geq 1$. Thus, condition (b) implies that μ_t does not decrease if we replace $\Pr[\Delta_t = -j \mid X^{(t)}]$ by $\delta^{-j} \Pr[\Delta_t = j \mid X^{(t)}]$ for every $j \geq 1$ and $\Pr[\Delta_t = 0 \mid X^{(t)}]$ by $1 - \sum_{j\geq 1}(1+\delta^{-j}) \Pr[\Delta_t = j \mid X^{(t)}]$. Hence,

$$\mu_t \leq 1 - \frac{(1-\delta^{-1/2})^2}{1 - \Pr[\Delta_t = 0 \mid X^{(t)}]} \sum_{j\geq 1} \Pr[\Delta_t = j \mid X^{(t)}] \, .$$

Now, again invoking condition (b),

$$\Pr[\Delta_t = j \mid X^{(t)}] \geq \frac{\Pr[\Delta_t = j \mid X^{(t)}] + \Pr[\Delta_t = -j \mid X^{(t)}]}{1+\delta^{-1}}$$

and since $\Pr[\Delta_t = 0 \mid X^{(t)}] = 1 - \sum_{j\neq 0} \Pr[\Delta_t = j \mid X^{(t)}]$ we have $\mu_t \leq 1 - 1/p(n)$.

4. Let $t \in \mathbb{N}$ and let $X^{(t)}$ be such that $g(X^{(t)}) \geq b(n)$, that is, $X^{(t)} \geq bn$. Then μ_t increases if we reduce $\Pr[\Delta_t = j]$ to zero for all $j \geq 1$ in combination with replacing $\Pr[\Delta_t = 0]$ by one. Thus,

$$\mu_t \leq 1 + \sum_{j\geq 1} \delta^j \Pr[\Delta_t = -j \mid X^{(t)}] \overset{(c)}{\leq} D \, .$$

Since conditions 1.−4. of the Drift Theorem hold,

$$\Pr[T \leq \delta^{1/3\,(b-a)\,n}] \leq \frac{(C+1)(1-\delta^{-1/2})^2}{1+\delta^{-1}} \, \delta^{-1/6\,(b-a)\,n} \, .$$

This is bounded from above by $2^{-\Omega(n)}$, hence the statement of this theorem follows. □

The theorem on negative drift (Theorem 4.9) in the last section is more recent and also more general than the previous theorem. We use Theorem 4.9 to prove Theorem 5.19 in the next chapter, yet using the previous theorem would be possible, too.

4.4. Dominance of Stochastic Processes

In the analysis of randomized search heuristics, there are situations where (potentially dependent) random variables can be bounded independently of each other (for example, in the proofs of Proposition 5.22 and Proposition 5.40 in the next chapter). The following theorem tells us that in these situation we can find independent random variables that dominate these random variables on a one-to-one basis.

Theorem 4.13. *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random variables over a finite space $\mathcal{S} \subseteq \mathbb{R}$. Let $n \in \mathbb{N}$, let $z \in \mathcal{S}$, and for each $x \in \mathcal{S}$ let p_x be a number in $[0,1]$ such that $\sum_{x\in S} p_x = 1$. Suppose that*

$$\Pr\left[X^{(t)} = x \,\middle|\, X^{(t-1)}, \ldots, X^{(0)}\right] \geq p_x \quad \text{if } x > z \text{ and}$$
$$\Pr\left[X^{(t)} = x \,\middle|\, X^{(t-1)}, \ldots, X^{(0)}\right] \leq p_x \quad \text{if } x < z$$

holds for all $t \in \{0, \ldots, n\}$ and $x \in \mathcal{S}$. Then, there exist a sequence of random variables $\{Y^{(t)}\}_{t\in\mathbb{N}}$ over $\mathcal{S}$ such that

1. $\Pr[Y^{(t)} = x] = p_x$ *holds for all* $x \in S$ *and* $t \in \{0, \ldots, n\}$*;*

2. $Y^{(t)} \leq X^{(t)}$ *holds for all* $t \in \{0, \ldots, n\}$*;*

3. *the random variables* $Y^{(0)}, \ldots, Y^{(n)}$ *are mutually independent; and*

4. *the distribution of* $Y^{(t)}$ *depends only on* $X^{(0)}, \ldots, X^{(t)}$ *for all* $t \in \{0, \ldots, n\}$*.*

Proof. Let $\mathcal{S}^+ := \{x \in \mathcal{S} \colon x > z\}$ and $\mathcal{S}^- := \{x \in \mathcal{S} \colon x < z\}$. We define the random variables $Y^{(t)}$ recursively.

For $t \in \{0, \ldots, n\}$ and $x \in \mathcal{S}$, let $q_x^{(t)} := \Pr[X^{(t)} = x \mid Y^{(t-1)}, \ldots, Y^{(0)}]$ and

$$\delta_x^{(t)} := \begin{cases} q_x^{(t)} - p_x & \text{if } x \in \mathcal{S}^+\,, \\ p_x - q_x^{(t)} & \text{if } x \in \mathcal{S}^-\,, \\ |q_z^{(t)} - p_z| & \text{if } x = z\,. \end{cases}$$

By the assumption, $\delta_x^{(t)}$ is positive for all $x \in \mathcal{S}$ since the distribution of $Y^{(s)}$ depends only on $X^{(0)}, \ldots, X^{(s)}$ for all $s < t$. For all $t \in \{0, \ldots, n\}$, let

$$P^{(t)} := \sum_{x \in \mathcal{S}^+} \delta_x^{(t)} \quad \text{and} \quad N^{(t)} := \sum_{x \in \mathcal{S}^-} \delta_x^{(t)}\,.$$

Furthermore, since $\sum_{x \in \mathcal{S}} q_x^{(t)} = 1 = \sum_{x \in \mathcal{S}} p_x$, we have $|P^{(t)} - N^{(t)}| = \delta_z^{(t)}$.

Let $t \in \{0, \ldots, n\}$ and $x, y \in \mathcal{S}$. In the case that $P^{(t)} \geq N^{(t)}$, let

$$s_{x,y}^{(t)} := \begin{cases} 1 & \text{if } x \in \mathcal{S}^- \cup \{z\} \text{ and } y = x\,, \\ \frac{p_x}{q_x^{(t)}} & \text{if } x \in \mathcal{S}^+ \text{ and } y = x\,, \\ \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{P^{(t)}\,q_x^{(t)}} & \text{if } x \in \mathcal{S}^+ \text{ and } y \in \mathcal{S}^- \cup \{z\}, \\ 0 & \text{otherwise.} \end{cases}$$

In the case that $P^{(t)} < N^{(t)}$, let

$$s_{x,y}^{(t)} := \begin{cases} 1 & \text{if } x \in \mathcal{S}^- \text{ and } y = x\,, \\ \frac{p_x}{q_x^{(t)}} & \text{if } x \in \mathcal{S}^+ \cup \{z\} \text{ and } y = x\,, \\ \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{N^{(t)}\,q_x^{(t)}} & \text{if } x \in \mathcal{S}^+ \cup \{z\} \text{ and } y \in \mathcal{S}^-, \\ 0 & \text{otherwise.} \end{cases}$$

In both cases, we choose $Y^{(t)}$ depending on $X^{(t)}$ and $Y^{(t-1)}, \ldots, Y^{(0)}$ such that

$$\Pr[Y^{(t)} = y \mid X^{(t)} = x, Y^{(t-1)}, \ldots, Y^{(0)}] = s_{x,y}^{(t)}\,.$$

Note that this distribution is well-defined. In the case that $P^{(t)} \geq N^{(t)}$ we have

$$\sum_{y \in \mathcal{S}} s_{x,y}^{(t)} = \begin{cases} 1 & \text{if } x \in \mathcal{S}^- \cup \{z\}\,, \\ \frac{p_x}{q_x^{(t)}} + \sum_{y \in \mathcal{S} \cup \{z\}} \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{P^{(t)}\,q_x^{(t)}} = 1 & \text{if } x \in \mathcal{S}^+\,. \end{cases}$$

In the case that $P^{(t)} < N^{(t)}$ we have

$$\sum_{y\in\mathcal{S}} s_{x,y}^{(t)} = \begin{cases} 1 & \text{if } x \in \mathcal{S}^-\,, \\ \frac{p_x}{q_x^{(t)}} + \sum_{y\in\mathcal{S}^-} \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{N^{(t)}\,q_x^{(t)}} = 1 & \text{if } x \in \mathcal{S}^+ \cup \{z\}\,. \end{cases}$$

In both cases it holds that $Y^{(t)} \leq X^{(t)}$ and that the distribution of $Y^{(t)}$ depends only on $X^{(0)}, \ldots, X^{(t)}$. Furthermore, we have $\sum_{x\in\mathcal{S}} s_{x,y}^{(t)}\,q_x^{(t)} = p_y$ for all $y \in \mathcal{S}$. In the first case,

$$\sum_{x\in\mathcal{S}} s_{x,y}^{(t)}\,q_x^{(t)} = \begin{cases} \frac{p_y}{q_y^{(t)}}\,q_y^{(t)} = p_y & \text{if } y \in \mathcal{S}^+\,, \\ q_y^{(t)} + \sum_{x\in\mathcal{S}^+} \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{P^{(t)}\,q_x^{(t)}}\,q_x^{(t)} = p_y & \text{if } y \in \mathcal{S}^- \cup \{z\}\,. \end{cases}$$

In the second case,

$$\sum_{x\in\mathcal{S}} s_{x,y}^{(t)}\,q_x^{(t)} = \begin{cases} \frac{p_y}{q_y^{(t)}}\,q_y^{(t)} = p_y & \text{if } y \in \mathcal{S}^+ \cup \{z\}\,, \\ q_y^{(t)} + \sum_{x\in\mathcal{S}^+\cup\{z\}} \frac{\delta_x^{(t)}\,\delta_y^{(t)}}{N^{(t)}\,q_x^{(t)}}\,q_x^{(t)} = p_y & \text{if } y \in \mathcal{S}^-\,. \end{cases}$$

Summing up, we get

$$\Pr[Y^{(t)} = y \mid Y^{(t-1)}, \ldots, Y^{(0)}] = \sum_{x\in\mathcal{S}} s_{x,y}^{(t)}\,q_x^{(t)} = p_y\,.$$

Hence, the $Y^{(t)}$'s are mutually independent with $\Pr[Y^{(t)} = y] = p_y$. $\square$

The previous theorem implies the following corollary which is used in the proof of Theorem 4.11.

Corollary 4.14. *Let $\{X^{(t)}\}_{t\in\mathbb{N}}$ be a sequence of random binary variables over $\{0,1\}$. Furthermore, let $n \in \mathbb{N}$ and $p \in [0,1]$.*

Suppose

$$\Pr\left[X^{(t)} = 1 \,\middle|\, X^{(t-1)}, \ldots, X^{(0)}\right] \geq p\,.$$

holds for all $t \in \{0, \ldots, n\}$. Then, there exist a sequence of random variables $\{Y^{(t)}\}_{t\in\mathbb{N}}$ over $\{0,1\}$ such that

1. $\Pr[Y^{(t)} = 1] = p$ *holds for all $t \in \{0, \ldots, n\}$;*
2. $Y^{(t)} \leq X^{(t)}$ *holds for all $t \in \{0, \ldots, n\}$;*
3. *the random variables $Y^{(0)}, \ldots, Y^{(n)}$ are mutually independent; and*
4. *the distribution of $Y^{(t)}$ depends only on $X^{(0)}, \ldots, X^{(t)}$ for all $t \in \{0, \ldots, n\}$.*

Indication of source. The content of this chapter contains parts that have been previously published in the *Proceedings of GECCO '08* (Happ, Johannsen, Klein, and Neumann (2008)) and the *Proceedings of GECCO '10* (Doerr, Johannsen, and Winzen (2010b)) and results that will appear in the *Proceedings of CEC '10* (Doerr, Johannsen, and Winzen (2010a)).

5
Evolutionary Computation in Combinatorial Optimization

This chapter has been written for the purpose of being a survey included in the book *Theory of Randomized Search Heuristics* (Auger and Doerr (2010)) and is presented as such.

Classical combinatorial optimization problems play a central role in the theoretical study of evolutionary algorithms. On the one hand, these problems are general enough to make meaningful comparisons among different evolutionary algorithms. On the other hand, combinatorial optimization problems have enough structural properties to make the theoretical analysis of such algorithms possible.

This chapter gives a concise overview over the runtime analysis of the (1+1) Evolutionary Algorithm (EA) on polynomially solvable problems in combinatorial optimization. We conduct these analyses for the minimum spanning tree problem in the context of maximum weight bases of matroids; for the single-source and all-pair shortest path problem; for the maximum matching problem; and for the problem of finding an Euler tour.

An extensive discussion of these problems and the corresponding problem-specific polynomial time algorithms can be found in Mehlhorn and Sanders (2009) and in Cormen, Leiserson, Rivest, and Stein (2001), for example.

We have selected these problems for two main reasons. First, they are among the most studied combinatorial problems in the theory of evolutionary algorithms. Second, these problems are particularly suited to demonstrate typical techniques used in the runtime analysis of evolutionary algorithms. For example, we will present the concepts of drift analysis; typical runs; dominance of stochastic processes; and large deviation bounds.

5.1. The Basic Combinatorial (1+1) Evolutionary Algorithm

A type of problem that occurs in classical combinatorics is to find a specific subgraph in a given graph (we consider all graphs to be finite, simple, and undirected unless explicitly stated otherwise). In particular, such problems assign an objective value to each subgraph and ask to maximize or minimize this value over all subgraphs of a given class. Prominent examples which we discuss in this chapter are the problems of finding a minimum spanning tree (MST), a single-source shortest path tree (SSSP), or a maximum matching (MM) in an edge-weighted graph.

Given a graph $G = (V, E)$, we identify a subgraph F of G with its edge set, that is, with slight abuse of notation we let F be an element of the power set 2^E of E. A subgraph optimization problem on G then consists of the space of *feasible* subgraphs $\mathcal{F} \subseteq 2^E$ and a partial order $\succeq$ on $\mathcal{F}$. A subgraph $F_{\text{opt}} \in \mathcal{F}$ is *optimal* if $F_{\text{opt}} \succeq F$ for all $F \in \mathcal{F}$. For all problems we consider in this chapter, there exists such an optimum.

Often, we define $\succeq$ implicitly by an objective function $f\colon 2^E \to \mathbb{R}$. Then, $F \succeq H$ if and only if $f(F) \geq f(H)$ in case f is to be maximized and $F \succeq H$ if and only if $f(F) \leq f(H)$ in case f is to be minimized. In both cases, all optimal solutions have the same objective value f_{opt}. Such an objective function may be given by edge-weights $w\colon E \to \mathbb{R}^+$. In this case, the objective value of a subgraph $F \subseteq E$ is the cumulative weight of its edges $w(F) := \sum_{e \in F} w(e)$. As first example of a subgraph optimization problem, we consider the MST problem.

Problem 5.1 (The Minimum Spanning Tree Problem).
Let $G = (V, E)$ be a connected graph and $w\colon E \to \mathbb{R}^+$ a weight function on E. The minimum spanning tree (MST) problem asks for a spanning tree $F \subseteq E$ that minimizes $w(F) := \sum_{e \in F} w(e)$.

For the MST problem, the space of feasible subgraphs $\mathcal{F}$ contains all spanning trees of G. For two spanning trees F and H, we let $F \succeq H$ (that is, F is "better" than H) if and only if $w(F) \leq w(H)$.

We are interested in how the basic combinatorial (1+1) EA solves subgraph optimization problems like the MST problem.

Algorithm 5.2 (The Basic Combinatorial (1+1) EA).
Let E be a finite set. Furthermore, let $\mathcal{F} \subseteq 2^E$ be a set of feasible search points, let $\succeq$ be a partial order relation on $\mathcal{F}$, and let $x^{(0)} \in \mathcal{F}$ be an initial search point. The basic combinatorial (1+1) EA corresponding to the tuple $(E, \mathcal{F}, \succeq, x^{(0)})$ iteratively generates a sequence of search points $\{x^{(t)}\}_{t \in \mathbb{N}}$ in $\mathcal{F}$ by the following procedure.

For all $t \in \mathbb{N}$ with $t \geq 1$, the basic combinatorial (1+1) EA generates a random candidate search point $y^{(t)} \subseteq E$ such that $y^{(t)}$ differs from $x^{(t-1)}$ in each edge $e \in E$ with probability $1/|E|$. In other words, $\Pr[e \in x^{(t-1)} \triangle y^{(t)}] = 1/|E|$ independently for all $e \in E$ (where $\triangle$ is the symmetric difference between two sets). Afterwards,

$$x^{(t)} := \begin{cases} y^{(t)} & \text{if } y^{(t)} \in \mathcal{F} \text{ and } y^{(t)} \succeq x^{(t-1)}\,, \\ x^{(t-1)} & \text{otherwise.} \end{cases}$$

In particular, we are interested in the *optimization time* of the basic combinatorial (1+1) EA. This is the random variable T that describes the first point in time $t \in \mathbb{N}$ for which $x^{(t)}$ is optimal.

Clearly, the basic combinatorial (1+1) EA is not only applicable to subgraph optimization problems but to all combinatorial problems that can be formulated in terms of a space $\mathcal{F}$ of feasible subsets over a ground set E and a partial order relation $\succeq$ that represents an optimization objective.

We conclude this section with a theorem on the optimization time of the (1+1) EA for the MST problem.

Algorithm 5.3 ((1+1) EA$_{\text{MST}}$). *The basic combinatorial (1+1) EA for the minimum spanning tree problem, the (1+1) EA$_{\text{MST}}$, is an instance of Algorithm 5.2.*

Let $G = (V, E)$ be a graph and $w\colon E \to \mathbb{R}^+$ a weight function on E. Then, the space of feasible search points $\mathcal{F}$ consists of all spanning trees of G and $F \succeq H$ holds for $F, H \in \mathcal{F}$ if $w(F) \leq w(H)$ holds.

The following result follows from the analysis in Neumann and Wegener (2007). We prove it in the more general setting of maximal weight bases in the following sections (the upper bound in Theorem 5.10 and the lower bound in Theorem 5.14). In these results we also specify leading constants of the given bounds and drop the restriction to integral edge weights.

Theorem 5.4 (Neumann and Wegener (2007)). *Let $G = (V, E)$ be a connected graph and let $w\colon E \to \mathbb{N}^+$ be a weight function on E with maximum weight $w_{\max}$.*

The expected optimization time of the (1+1) EA$_{\text{MST}}$ on (G, w) starting with an arbitrary initial search point $x^{(0)}$ is $\mathcal{O}\big(|E|^2 (\ln |V| + \ln w_{\max})\big)$

Furthermore, for every $n \in \mathbb{N}$ there exists a connected graph $G = (V, E)$ on n vertices and a weight function $w\colon E \to \mathbb{N}^+$ such that expected optimization time of the (1+1) EA$_{\text{MST}}$ (Algorithm 5.3) on (G, w) with the initial search point chosen uniformly at random from all spanning trees on G is $\Theta\big(|E|^2 \ln |V|\big)$.

Neumann and Wegener (2005) also analyzed multi-objective evolutionary algorithms for the MST problem. They showed for two such algorithms a superior expected optimization time of $\mathcal{O}\big(|V|\,|E|\,(\ln |V| + \ln w_{\max})\big)$.

5.2. Matroids — The Realm of the Greedy Algorithm

As a consequence of Theorem 5.4 from the previous section, the (1+1) EA$_{\text{MST}}$ solves the MST problem in expected time $\mathcal{O}(|E|^2 \ln |V|)$ for polynomially bounded integer weights. In comparison, the two non-evolutionary greedy algorithms by Kruskal (1956) and by Jarník and Prim (Jarník (1930); Prim (1957)) solve this problem in times $\mathcal{O}(|E| \ln |V|)$ and $\mathcal{O}(|E| + |V| \ln |V|)$, respectively.

As hill-climber strategies, greedy algorithms are one of the most basic tools in black-box optimization. One of the most general problems known to be tractable for greedy algorithms is the problem of finding a maximum weight basis of a matroid.

Definition 5.5 (Matroid). *Let E be a finite set and $\mathcal{F} \subseteq 2^E$ be a set of subsets of E which are called independent sets of M. Then the pair $M = (E, \mathcal{F})$ is called an independence system if the two conditions*

(i) $\emptyset \in \mathcal{F}$ *and*

(ii) $\forall\, F \in \mathcal{F},\ H \subseteq F\colon H \in \mathcal{F}$

hold and is called a matroid if in addition the condition

(iii) $\forall\, F, H \in \mathcal{F},\ |H| < |F|\colon \exists\, e \in F \setminus H$ *such that* $H \cup \{e\} \in \mathcal{F}$

holds. An inclusion maximal set $B \in \mathcal{F}$ of an independence system is called a basis of M.

Condition (iii) in the previous definition implies that all bases of a matroid M have the same size $r(M)$, the *rank* of M. Another consequence of condition (iii) is the following *exchange property* (confer Reichel and Skutella (2007)).

Lemma 5.6. *Let $M = (E, \mathcal{F})$ be a matroid, F an independent set of M, and B a basis of M. Furthermore, let $H_F := F \setminus B$ and $H_B := B \setminus F$. Then there exist an injective map $\varphi\colon H_F \to H_B$ such that*

(i) $F \cup \{\varphi(f)\} \setminus \{f\}$ *is independent for all* $f \in H_F$ *and*

(ii) $F \cup \{b\}$ *is independent for all* $b \in H_B \setminus \varphi(H_F)$.

Proof. If F is not a basis then by condition (iii) of Definition 5.5 we can successively add elements from H_B to F until F is a basis. These elements then satisfy condition (ii) of the statement. Thus, we may suppose F is a basis. In this case $|F| = |B|$ and we only have to verify condition (i) of the statement.

Consider the bipartite graph on the two partitions H_F and H_B such that for $f \in H_F$ and $b \in H_B$ the pair (f, b) is an edge if $F \cup \{b\} \setminus \{f\}$ is independent. Suppose for all $H \subseteq H_F$ the inequality $|N(H)| \geq |H|$ holds where $N(H)$ is the neighborhood of H in B. Then by the Theorem of Hall (see, e. g., Diestel (2005)) there exists a bipartite matching that covers H_F which defines the function φ satisfying condition (i) of the statement.

For proof by contradiction, let us assume that there exists a subset H of H_B such that $|N(H)| < |H|$. Then, by condition (iii) of Definition 5.5, there exists a $b \in H$ such that $(F \setminus H_F) \cup N(H) \cup \{b\}$ is independent. Again by (iii), we find $f_1, \ldots, f_{|F_H|-|N(H)|-1}$ in $F_H \setminus N(H)$ such that the set $F' = (F \setminus F_H) \cup N(H) \cup \{b, f_1, \ldots, f_{|F_H|-|N(H)|-1}\}$ remains independent. Since $|F'| = |F|$, there exists an element $f \in F_H \setminus N(H)$ that is not in F'. But then (f, h) is an edge — a contradiction to $f \notin N(H)$. □

For a graph $G = (V, E)$, let $\mathcal{F}$ be the set of edge sets $F \subseteq E$ such that the subgraph (V, F) is acyclic. Then $M_G = (E, \mathcal{F})$ is a matroid, called the *graph matroid* of G. In M_G, the spanning trees of G correspond to the bases of M_G which have all size $n - 1$. Thus, the *maximum*[1] spanning tree problem is equivalent to the problem of finding a maximum weight basis of the corresponding graph matroid.

[1]The minimum spanning tree and the maximum spanning tree problems can be easily transformed into each other by replacing $w(e)$ by $w_{\max} + 1 - w(e)$ for all edges $e \in E$.

Problem 5.7 (The Maximum Weight Basis Problem).
Let $M = (E, \mathcal{F})$ be an independence system and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. The maximum weight basis problem asks for a set $F \in \mathcal{F}$ that maximizes $w(F) := \sum_{e \in F} w(e)$. Note that such a set is necessarily a basis.

Given an independence system $M = (E, \mathcal{F})$, the greedy algorithm for the maximum weight basis problem starts with the empty set. It then iteratively adds the element of largest weight that does not violate the independence of the current set. This greedy algorithm finds a maximum weight basis of M for every weight function $w\colon E \to \mathbb{R}^+$ if and only if M is a matroid (see, e. g., Cormen et al. (2001)).

For a matroid $M = (E, \mathcal{F})$ with known weights, the greedy algorithm can be implemented by first sorting the elements in E and then checking independence in decreasing order. The computation time of this procedure is $\mathcal{O}(|E| \ln |E| + f(|E|))$ where $f(k)$ is the time needed to check k sets for independence.

Unlike the greedy algorithm, the basic combinatorial (1+1) EA solves the maximum weight basis problem also for non-matroidal independence systems. Since it is more generic, we do not expect it to outperform the greedy algorithm if the problem instance is a matroid. Still, in the next section we show an upper bound of $\mathcal{O}(|E|^3 \ln |E|)$ on the expected optimization time of the basic combinatorial (1+1) EA which reduces to $\mathcal{O}(|E|^2 \ln |E|)$ for polynomially bounded integer weights.

5.3. Multiplicative Drift Analysis

We next show an upper bound on the optimization time of the (1+1) EA on the maximum weight basis problem.

Algorithm 5.8 ((1+1) EA$_{\mathrm{MWB}}$). *The basic combinatorial (1+1) EA for the maximum weight basis problem, the (1+1) EA$_{\mathrm{MWB}}$, is an instance of Algorithm 5.2.*

Let $M = (E, \mathcal{F})$ be a matroid and $w\colon E \to \mathbb{R}^+$ a weight function on E. Then, $\mathcal{F}$ is the space of feasible search points and $F \succeq H$ holds for $F, H \in \mathcal{F}$ if $w(F) \geq w(H)$ holds.

We study the average weight increase in each iteration of this algorithm, confer Reichel and Skutella (2007).

Proposition 5.9. *Let $M = (E, \mathcal{F})$ be a matroid and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Let w_{opt} be the weight of a maximum weight basis of M. Furthermore, let $\{x^{(t)}\}_{t \in \mathbb{N}}$ be the sequence of search points generated by the (1+1) EA$_{\mathrm{MWB}}$ on (M, w). Then,*

$$\mathrm{E}\big[w(x^{(t+1)}) - w(x^{(t)}) \,\big|\, w(x^{(t)})\big] \geq \frac{w_{\mathrm{opt}} - w(x^{(t)})}{\mathrm{e}\,|E|^2},$$

for all $t \in \mathbb{N}$ where $\mathrm{e} = 2.718\ldots$ is the Euler constant.

Proof. Let $t \in \mathbb{N}$ and $m := |E|$. Let B be a maximum weight basis ($w(B) = w_{\mathrm{opt}}$) and let $F := x^{(t)}$ with $t \in \mathbb{N}$. Furthermore, let $\varphi\colon H_F \to H_B$ with $H_F := B \setminus F$ and $H_B := F \setminus B$ be the injective function that is provided by Lemma 5.6.

For all $f \in H_F$ we define the indicator variable I_f by

$$I_f := \begin{cases} 1 & \text{if } x^{(t+1)} = x^{(t)} \cup \{\varphi(f)\} \setminus \{f\}, \\ 0 & \text{otherwise,} \end{cases}$$

and for all $b \in H_B \setminus \varphi(H_F)$ the indicator variable I_b by

$$I_b := \begin{cases} 1 & \text{if } x^{(t+1)} = x^{(t)} \cup \{b\}, \\ 0 & \text{otherwise.} \end{cases}$$

Since the variables I_f with $f \in H_F$ and I_b with $b \in H_B \setminus \varphi(H_F)$ indicate disjoint events,

$$\sum_{f \in H_F} I_f + \sum_{b \in H_B \setminus \varphi(H_F)} I_b \le 1$$

holds. Moreover, according to the definition of the (1+1) EA, it holds that $w(x^{(t+1)})$ is at least as large as $w(x^{(t)})$. Thus,

$$w(x^{(t+1)}) - w(x^{(t)}) \ge \sum_{f \in H_F} \big(w(b) - w(f)\big)\, I_f + \sum_{b \in H_B \setminus \varphi(H_F)} w(b)\, I_b\,.$$

With this inequality at hand, we bound the expected weight increase. For all $f \in H_F$ and all $b \in H_B \setminus \varphi(H_F)$ let $p_f := \Pr[I_f = 1 \mid w(x^{(t)})]$ and $p_b := \Pr[I_b = 1 \mid w(x^{(t)})]$. Then,

$$\mathrm{E}\big[w(x^{(t+1)}) - w(x^{(t)}) \,\big|\, w(x^{(t)})\big] \ge \sum_{f \in H_F} \big(w(b) - w(f)\big)\, p_f + \sum_{b \in H_B \setminus \varphi(H_F)} w(b)\, p_b\,.$$

To conclude the proof of this proposition, it suffices to show that p_f and p_b are at least $p := \mathrm{e}^{-1} m^{-2}$ for all $f \in H_F$ with $w(f) \le w(\varphi(f))$ and for all $b \in H_B \setminus \varphi(H_F)$.

Let f be an element of H_F such that $w(f) \le w(\varphi(f))$. Then p_f is the probability of $x^{(t+1)} = x^{(t)} \cup \{\varphi(f)\} \setminus \{f\}$. That is, p_f is the probability of $x^{(t)} \triangle y^{(t+1)} = \{f, \varphi(f)\}$ where $y^{(t+1)}$ is the respective candidate search point of the (1+1) EA. Thus,

$$p_f = \Big(1 - \frac{1}{m}\Big)^{m-2} \frac{1}{m^2} \ge \frac{1}{\mathrm{e}\, m^2}\,.$$

Similarly, if $b \in H_B \setminus \varphi(H_F)$ then p_b is the probability of $x^{(t)} \triangle y^{(t+1)} = \{b\}$. Thus

$$p_b = \Big(1 - \frac{1}{m}\Big)^{m-1} \frac{1}{m} \ge \frac{1}{\mathrm{e}\, m}\,.$$

□

The previous proposition shows that on average the distance in weights towards an optimal search point decreases at least proportional to its current value in each iteration. In such a situation, the expected optimization time can be bounded from above using the theorem on multiplicative drift (Theorem 4.5) from the previous chapter.

Let $M = (E, \mathcal{F})$ be a matroid and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Then, the previous theorem allows us to bound the expected optimization time of the (1+1) $\mathrm{EA}_{\mathrm{MWB}}$ on a weighted matroid (M, w). Let w_{opt} be the weight of a maximum weight basis and let $g(F) := w_{\mathrm{opt}} - w(F)$ for all $F \in \mathcal{F}$ be a potential function on $\mathcal{F}$. Then the following statement directly follows from Proposition 5.9.

Theorem 5.10 (confer Reichel and Skutella (2007)). *Let $M = (E, \mathcal{F})$ be a matroid and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Let w_{opt} be the weight of a maximum weight basis and $w_{\text{2nd-opt}}$ be the maximum weight over all bases that do not have weight w_{opt}. Furthermore, let T be the optimization time of the (1+1) EA$_{\text{MWB}}$ (Algorithm 5.8) on (M, w). Then,*

$$\mathrm{E}[T \mid x^{(0)}] \le \mathrm{e}\,|E|^2 \left(1 + \ln \frac{w_{\text{opt}} - w(x^{(0)})}{w_{\text{opt}} - w_{\text{2nd-opt}}}\right) .$$

As a consequence of this theorem, the expected optimization time of (1+1) EA$_{\text{MWB}}$ on (M, w) is $\mathcal{O}(|E|^2 \ln |E|)$ for integer weights that polynomially bounded in $|E|$. For non-integer weights a weight-independent upper bound can be inferred from the following theorem.

Theorem 5.11 (Reichel and Skutella (2009)). *Let E be a finite index set, $\mathcal{F} \subseteq 2^E$ a search space, and $w\colon E \to \mathbb{N}^+$ a weight function on E. Then there exists a bounded weight function $\widetilde{w}\colon E \to \{0, \ldots, |E|^{|E|/2}\}$ on E such that the two corresponding sequences of search points $\{x^{(t)}\}_{t\in\mathbb{N}}$ and $\{\widetilde{x}^{(t)}\}_{t\in\mathbb{N}}$ generated by the basic combinatorial (1+1) EA (Algorithm 5.2) are the same for all choices of $x^{(0)}$.*

This theorem also holds for arbitrary weight functions $w\colon E \to \mathbb{R}^+$. Since the number of weights is finite, we can scale all weights by a very large constant until rounding to the next integer does not change the behavior of the (1+1) EA anymore. Then we apply the previous theorem. This implies that the expected optimization time of the (1+1) EA$_{\text{MWB}}$ on (M, w) with an arbitrary weight function $w\colon E \to \mathbb{R}^+$ is $\mathcal{O}(|E|^3 \ln |E|)$.

Theorem 5.12 (confer Reichel and Skutella (2009)). *Let $M = (E, \mathcal{F})$ be a matroid and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Furthermore, let T be the optimization time of the (1+1) EA$_{\text{MWB}}$ (Algorithm 5.8) on (M, w). Then for $|E| \ge 2$, independent of $x^{(0)} \in \mathcal{F}$,*

$$\mathrm{E}[T] \le \tfrac{\mathrm{e}}{2}\,|E|^3\bigl(1 + \ln|E|\bigr) .$$

5.4. Lower Bounds and Typical Runs

In the previous section we have proven upper bounds on the optimization time of the maximum weight basis problem. The following example will serve as instance for a lower bound. It can be perceived as a graph matroid corresponding to the graph with multiple edges depicted in Figure 5.1.

Definition 5.13. *For all $m \in \mathbb{N}$ with $m \ge 3$, let the matroid $M_m = (E, \mathcal{F})$ and the weight function $w_m\colon E \to \{1, 2, m, m+3\}$ be defined as follows.*

Let $k := \lfloor m^{1/3} \rfloor$ and $E := \{e_1, \ldots, e_k, f_1, \ldots, f_k, g, h_1, \ldots, h_{m-2k-1}\}$. A subset F of E is independent if for all $i \in \{1, \ldots, k\}$ it does not contain both, e_i and f_i, and at

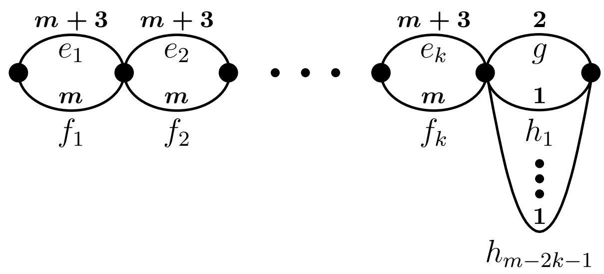

Figure 5.1. The above (multi-)graph is a chain of k pairs of parallel edges terminated by a multi-edge consisting of $m-2k$ parallel edges. Each pair has an edge of weight $m+3$ and one of weight m. Of the $m-2k$ parallel edges, one edge has weight 2 and the others have weight 1. The independent sets of the corresponding graph matroid are the cycle-free subgraphs of this graph, that is, an independent set contains at most one edge of each pair and at most one of the $m-2k$ parallel edges. For $k := \lfloor \sqrt[3]{m} \rfloor$, the expected time for the (1+1) EA to find the optimal search point for the minimum weight basis problem (which constitutes of the row of upper edges) starting with a random feasible set of edges is $\Omega(m^2 \ln m)$.

most one of the elements $g, h_1, \ldots, h_{m-2k-1}$. The weight function w_m is defined by

$$w_m(e) := \begin{cases} m+3 & \text{if } e = e_i \text{ with } i \in \{1, \ldots, k\}, \\ m & \text{if } e = f_i \text{ with } i \in \{1, \ldots, k\}, \\ 2 & \text{if } e = g, \\ 1 & \text{if } e = h_i \text{ with } i \in \{1, \ldots, m-2k-1\}. \end{cases}$$

For M_m with w_m, we consider a *typical run*. This means, we show that the sequence $\{x^{(t)}\}_{t \in \mathbb{N}}$ generated by the (1+1) EA has certain properties with a probability that is bounded from below by a positive constant. These properties then imply an optimum is not found within $\Omega(|E|^2 \ln |E|)$ iterations.

Theorem 5.14 (confer Neumann and Wegener (2007)). *Let $m \in \mathbb{N}$ with $m \geq 3$. Furthermore, let M_m be the matroid and $w_m \colon E \to \mathbb{R}^+$ be the weight function from Definition 5.13. Then the expected optimization time of the (1+1) EA_{MWB} (Algorithm 5.8) on (M_m, w_m) with $x^{(0)}$ chosen uniformly at random from $\mathcal{F}$ is $\Omega(|E|^2 \ln |E|)$.*

Proof. Since we prove an asymptotic result, we may suppose that in the following m is sufficiently large. Let $x^{(0)}$ be chosen uniformly at random from $\mathcal{F}$. Let $\{x^{(t)}\}_{t \in \mathbb{N}}$ be the sequence of search points generated by the (1+1) EA_{MWB} on (M_m, w_m). Furthermore, let T be the random variable that denotes the earliest point in time $t \in \mathbb{N}$ for which $x^{(t)}$ is the maximum weight basis $\{e_1, \ldots, e_k, g\}$ of M_m.

Let $I := \{1, \ldots, k\}$. For $i \in I$ and $t \in \mathbb{N}$ we say that at time t position i is *free* if neither e_i nor f_i is in $x^{(t)}$ and *occupied* otherwise. We call it *critical* if $f_i \in x^{(s)}$ for all $s \leq t$. Clearly, a critical position is never free and $T > t$ if there exist any critical positions at time t.

For the remainder of the proof, we neglect whether or not any one of the elements $g, h_1, \ldots, h_{m-2k-1}$ is in $x^{(t)}$. We can do this, since these events have no influence on the number of occupied and critical positions.

Since $x^{(0)}$ is chosen uniformly at random, in expectation $k/3$ positions are critical at time $t_0 := 0$. Consider the event that at time t_0 there are at least $k/4$ positions that are critical. Markov's inequality (Theorem 4.1) guarantee this event with a probability bounded from below by a positive constant.

First, we show that there are no free positions at time $t_1 := \lfloor 2\,\mathrm{e}\, m \ln m \rfloor$ with a probability that tends to one as m tends to infinity. .

Let $i \in I$ be a position that is free at time t. Let p be the probability that i is occupied at time $t+1$. Then p is at least the probability that $x^{(t)} \triangle y^{(t+1)} = \{e_i\}$. Thus,

$$p \geq \left(1 - \frac{1}{m}\right)^{m-1} \frac{1}{m} \geq \frac{1}{\mathrm{e}\, m}\,.$$

Now, let the potential function $g\colon \mathcal{F} \to \mathbb{N}$ measure the number of free positions in an independent set. Since there are $g(x^{(t)})$ free positions at time t, the expected decrease in g between t and $t+1$ is at least $g(x^{(t)})/em$. By Theorem 4.5, the expected time until no free positions remain is at most $em \ln m$. Thus, by the Chernoff bounds (Theorem 4.2), the probability that there are no free positions after t_1 iterations tends to one as m tends to infinity.

Next, we show that at time t_1 all positions that were critical at time t_0 are still critical with a probability that tends to one as m tends to infinity.

Let $i \in I$ be critical at time t. The only way for i to become non-critical at time $t+1$ is if $x^{(t)} \triangle y^{(t+1)}$ contains f_i and at least one e_j with $j \in I$ or if it contains f_j with $j \in I \setminus \{i\}$. This happens with probability at most $2\,k\,m^{-2}$. Thus, by the union bound, the probability that any position becomes non-critical at any point in time prior to t_1 is at most $4\,\mathrm{e}\,k^2\,m^{-1} \ln m$ which tends to zero as m tends to infinity.

Finally, we show that with a probability bounded from below by a positive constant, at time $t_2 := \lfloor \frac{1}{3} m^2 \, \ln m \rfloor$ there still exists a critical position.

We have already seen that with probabilities that tend to one as m tends to infinity, the events (i) that there are at least $k/4$ critical positions at time t_1 and (ii) that there are no free positions at time t_1 happen. Thus, by the union bound, (i) and (ii) occur simultaneously with a probability that is bounded away from zero by a positive constant if m is sufficiently large.

Suppose at time t_1 there exists at least $k/4$ critical and no free positions. Then, between the times t_1 and t_2 a critical position $i \in I$ can only become non-critical if the event "$x^{(t)} \triangle y^{(t+1)} \supseteq \{f_i, e_i\}$" occurs. This happens with probability at most m^{-2}. Therefore, with probability at least

$$\left(1 - \frac{1}{m^2}\right)^{\frac{1}{3} m^2 \, \ln m} \sim \frac{1}{m^{1/3}}$$

position i remains critical until time t_2.

Since the events "$x^{(t)} \triangle y^{(t+1)} \supseteq \{f_i, e_i\}$" and "$x^{(s)} \triangle y^{(s+1)} \supseteq \{f_j, e_j\}$" are mutually independent for $i \neq j$ and all combinations of s and t, the probability that there exist no critical positions at time t_2 is at most $(1 - m^{-1/3})^k \sim 1/\mathrm{e}$. Thus, with a probability

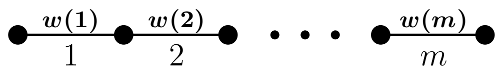

Figure 5.2. Maximizing a linear pseudo-Boolean function $f(x) = \sum_{i=1}^{m} w(i)\, x_i$ with strictly positive weight $w(1), \ldots, w(m)$ is equivalent to finding a maximum spanning tree on the path of length m with edges $1, \ldots, m$ and weights $w(1), \ldots, w(m)$. For this problem, the (1+1) EA finds an optimal search point in expected time $\Theta(m \ln m)$, independent of the weights $w(1), \ldots, w(m)$, see Droste et al. (2002).

bounded from below by a positive constant, the optimization time is at least t_2 which concludes the proof. □

With the previous theorem, we have seen that for matroids with an integer weight function polynomially bounded in $|E|$ the $\mathcal{O}(|E|^2 \ln |E|)$ bound on the expected optimization time of the (1+1) EA_{MWB} from the previous section is tight. However, for unbounded and real weights we only know the weight dependent bound from Theorem 5.10 and the general bound of $\mathcal{O}(|E|^3 \ln |E|)$ from Corollary 5.12.

It is a central open problem to close the gap between these two bounds. For the special case of linear pseudo-Boolean functions this question can be answered positively (see Figure 5.2). This leads us to conjecture that there exists a weight-independent upper bound of $\mathcal{O}(|E|^2 \ln |E|)$ on the expected optimization time of the (1+1) EA_{MWB}.

5.5. A Hard Problem for the (1+1) Evolutionary Algorithm

In the previous sections we have seen that the basic combinatorial (1+1) EA solves the MST and the SSSP problems in expected polynomial time if the weights are integral and polynomially bounded. A third classical problem which is known to be polynomially solvable is the *maximum matching problem*. Micali and Vazirani (1980) showed that this problem can be solved in time $\mathcal{O}(|V|^{1/2}\, |E|)$.

Problem 5.15 (The Maximum Matching Problem).
Let $G = (V, E)$ be a (multi-)graph. The maximum matching problem asks for a maximum set of vertex disjoint edges.

The basic combinatorial (1+1) EA for this problem is defined as follows.

Algorithm 5.16 ((1+1) EA_{MM}). *The basic combinatorial (1+1) EA for the maximum matching problem, the (1+1) EA_{MM}, is an instance of Algorithm 5.2. Let $G = (V, E)$ be a graph. Then, the space of feasible search points $\mathcal{F}$ consists of all matchings on G and $F \succeq H$ holds for $F, H \in \mathcal{F}$ if $|F| \geq |H|$ holds.*

This algorithm was studied extensively by Giel and Wegener (2003, 2006) and later by Oliveto, He, and Yao (2008) in the context of processes with negative drift. For example, Giel and Wegener (2003) showed that the expected optimization time (1+1) EA_{MM} on paths is polynomially bounded in $|E|$.

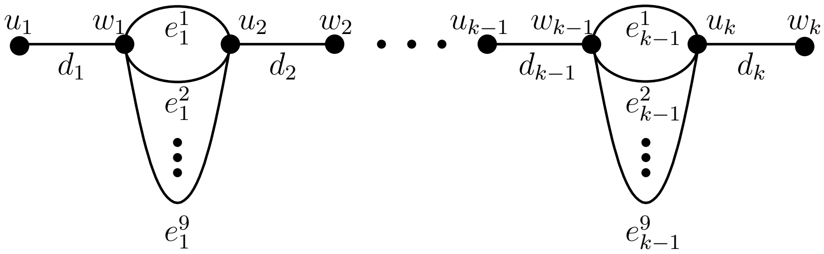

Figure 5.3. The multi-edged path P_k on $2k$ vertices and $10k-9$ edges for $k \geq 3$. Every *odd* edge d_i is a single edge and every *even* edge e_i^j is one of nine parallel edges. In an iteration of the (1+1) EA for the maximum matching problem it is roughly nine times more likely to replace a specific odd edge by an arbitrary neighboring even edge than vice versa. Because of this, augmenting paths on even edges have a strong tendency to grow.

Theorem 5.17 (Giel and Wegener (2003)). *Let $m \in \mathbb{N}$. The expected optimization time of the (1+1) EA_{MM} (Algorithm 5.16) on a path of m edges is $\mathcal{O}(m^4)$ independent of the initial search point $x^{(0)}$.*

Moreover, Giel and Wegener (2003) have shown that the (1+1) $\mathrm{EA}_{\mathrm{MM}}$ is a PRAS for the maximum matching problem. However, they also studied an instance of a bipartite graph due to Sasaki and Hajek (1988) for which the (1+1) $\mathrm{EA}_{\mathrm{MM}}$ has exponential expected optimization time.

In this section we show such an exponential lower bound on the expected optimization time for the multi-graph[2] in Figure 5.3.

Definition 5.18. *Let $k \in \mathbb{N}_{\geq 3}$ and let P_k be the multi-edged path on the set of vertices $\{u_1, w_1, \ldots, u_k, w_k\}$ such that there is a single edge d_i between u_i and w_i for all $1 \leq i \leq k$ and nine parallel edges $e_i^1, \ldots, e_i^9$ between w_i and u_{i+1} for all $1 \leq i \leq k-1$. Then, P_k has $m = 10k - 9$ edges in total. For simplicity, we call an edge d_i odd and an edge e_i^j even.*

The choice of this graph over the original graph in Giel and Wegener (2003) strongly simplifies the analysis while preserving the main proof ideas. Furthermore, it exposes a difficulty that exists for the (1+1) $\mathrm{EA}_{\mathrm{MM}}$ but not for problem-specific algorithms: We can adapt any problem-specific algorithm for simple graphs to an equally efficient algorithm for multi-graphs. The algorithm simply performs an initialization step where all multi-edges are replaced by single edges. Then, a maximum matching on this modified graph is also a maximum matching on the original multi-graph. In comparison, the (1+1) $\mathrm{EA}_{\mathrm{MM}}$ solves the maximum matching problem on paths in expected polynomial time. But, if we replace every second edge of a path of odd length by a multi-edge with nine parallel edges, then the expected optimization time of the (1+1) $\mathrm{EA}_{\mathrm{MM}}$ becomes exponential.

[2]A multi-graph is a graph with parallel edges.

Theorem 5.19 (confer Giel and Wegener (2003)). *Let $k \in \mathbb{N}_{\geq 3}$ and let P_k be the multi-edged path on $m = 10k - 9$ edges defined in Definition 5.18. Then, the expected optimization time of the (1+1) $\mathrm{EA_{MM}}$ (Algorithm 5.16) on P_k starting with a non-optimal initial search point is $2^{\Omega(m)}$.*

Before proving this theorem, we make a couple of preliminary observations. In particular, we first show three propositions that then lead to the proof of the previous theorem.

Let $P := P_k$ with $k \geq 3$ and $m = 10k - 9$. The set of odd edges of P forms the only perfect matching M^* on P. Thus, M^* is the unique maximum matching on P. It is of size k.

Let $\{x^{(t)}\}_{t\in\mathbb{N}}$ be the sequence of search points generated by the (1+1) $\mathrm{EA_{MM}}$ for the maximum matching problem on P initialized with an arbitrary matching $x^{(0)}$ other than M^*. Furthermore, let T^* be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $x^{(t)} = M^*$, that is the optimization time of the (1+1) $\mathrm{EA_{MM}}$.

We call a matching *almost perfect* if it contains $k - 1$ edges. For example, let M° be the almost perfect matching $M^* \triangle \{d_1, e_1^1, d_2\}$. We first show that we only loose a factor of $\Theta(1/m^3)$ on any lower bound for the expected optimization time of the (1+1) $\mathrm{EA_{MM}}$ if we condition on the event that $x^{(0)} = M^\circ$.

Proposition 5.20.

$$\mathrm{E}[T^* \mid x^{(0)} \neq M^*] \geq \frac{1}{2m^3}\,\mathrm{E}[T^* \mid x^{(0)} = M^\circ]\,.$$

Proof. Let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $x^{(t)}$ is either M° or M^*. We show that

$$\Pr[x^{(T)} = M^\circ] \geq \frac{1}{2m^3}\,. \tag{5.5.1}$$

This is sufficient as the statement then follows from the law of total expectation and from

$$\mathrm{E}[T^* \mid x^{(T)} = M^\circ] = \mathrm{E}[T \mid x^{(T)} = M^\circ] + \mathrm{E}[T^* \mid x^{(0)} = M^\circ]\,.$$

Let $t \in \mathbb{N}$ with $x^{(t)} \neq M^*$. Then, $|x^{(t)}| \leq k - 1$ and $y^{(t+1)}$ is accepted by the (1+1) $\mathrm{EA_{MM}}$ if $y^{(t+1)} = M^\circ$. Since $x^{(t)} \triangle M^\circ$ and $x^{(t)} \triangle M^*$ differ exactly by the three edges d_1, e_1^1, d_2, it holds that

$$\Pr[x^{(t+1)} = M^\circ] \geq \frac{1}{m^3}\,\Pr[x^{(t+1)} = M^*]$$

and inequality (5.5.1) follows from $\Pr[x^{(T)} = M^\circ] + \Pr[x^{(T)} = M^*] = 1$. □

Justified by this result, from now on we condition on the event that $x^{(0)} = M^\circ$. Since M° is an almost perfect matching, for $t < T^*$ all matchings $x^{(t)}$ are also almost perfect.

Every almost perfect matching M defines a unique *augmenting sub-path* S of P. More precisely, there exists two indices $a, b \in \{1, \ldots, k\}$ with $a \leq b$ such that the path $S := (u_a, d_a, w_a, e_a, \ldots, e_{b-1}, u_b, d_b, w_b)$ consist of all odd edges $d_a, \ldots, d_b$ that are unmatched and all even edges $e_a, \ldots, e_{b-1}$ that are matched. Let $\ell(S) := b - a$ be the number of even edges in S. Thus, S is of odd length $2\,\ell(s)+1$, since S always starts and ends with an odd edge (where $e_a = e_b$ is possible). For example, the alternating path in M° is $S^+ := (u_1, d_1, w_1, e_1^1, u_2, d_2, w_2)$ with $\ell(S^+) = 1$.

With slight abuse of notation, we associate with the matching M^* the *empty augmenting path* $S^* := (v_0)$ such that $\ell(S^*) = 0$. Note that in general an augmenting path S with $\ell(S) = 0$ can also correspond to an almost perfect matching and contain a single unmatched odd edge. We are not interested in this distinction. Instead, we let T_0 be the random variable that denotes the first point in time t such that $\ell(S^{(t)}) = 0$ and use it as a suitable lower bound of T^*.

Proposition 5.21.

$$\mathrm{E}[T^* \mid x^{(0)} = M^\circ] \geq \mathrm{E}[T_0 \mid \ell(S^{(0)}) = 1]\,.$$

The remainder of this section is devoted to proving a lower bound on $\mathrm{E}[T_0]$. On the one hand, we start with $\ell(S^{(0)}) = 1$ and it is quite likely that $\ell(S^{(t)})$ reaches zero for early points in time. On the other hand, we will see that the drift of $\ell(S^{(t)})$ is positive and thus $\ell(S^{(t)}$ has the tendency to increase up to $k-1$. The following proposition shows that the latter event happens with a probability that is bounded away from zero by a constant.

Proposition 5.22. *There exists a constant $\delta > 0$ such that*

$$\mathrm{E}[T_0 \mid \ell(S^{(0)}) = 1] \geq \delta\, \mathrm{E}[T_0 \mid \ell(S^{(0)}) = k-1]\,.$$

Proof. Let $\ell(S^{(0)}) = 1$ and let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $\ell(S^{(t)})$ is either 0 or $k-1$. Similar to Proposition 5.20, we show that there exists a constant $\delta > 0$ such that

$$\Pr[\ell(S^{(T)}) = k] \geq \delta\,.$$

Let $t < T$ and let $S^{(t)} = (v_{a(t)}, \ldots, w_{b(t)})$ be the augmenting path corresponding to the almost perfect matching $x^{(t)}$.

Since $t < T$, we have $a(t) > 1$ or $b(t) < k$, say $b(t) < k$ without loss of generality. Then, adding any of the edges $e^1_{b(t)}, \ldots, e^9_{b(t)}$ to $x^{(t)}$ and removing $d_{b(t)+1}$ increases the length of $S^{(t)}$ by two (and thus $\ell(S^{(t)})$ by one). Independent of $S^{(t)}, \ldots, S^{(0)}$, this increase happens with probability at least $9\,(1-1/m)^{m-2}\,m^{-2}$. Thus,

$$\Pr\left[\ell(S^{(t+1)}) - \ell(S^{(t)}) = 1 \,\middle|\, S^{(t)}, \ldots, S^{(0)}\right] \geq \frac{3}{m^2}\,. \tag{5.5.2}$$

On the other hand, shortening $S^{(t)}$ by two is not equally likely. We either have to add $d_{a(t)}$ to $x^{(t)}$ and remove $e_{a(t)}$ in return or have to add $d_{b(t)}$ to $x^{(t)}$ and remove $e_{b(t)-1}$. Thus, again independent of $S^{(t)}, \ldots, S^{(0)}$,

$$\Pr\left[\ell(S^{(t+1)}) - \ell(S^{(t)}) = -1 \,\middle|\, S^{(t)}, \ldots, S^{(0)}\right] \leq \frac{2}{m^2}\,. \tag{5.5.3}$$

More generally, to shorten $S^{(t)}$ by at least $2j$ edges, we have to change $x^{(t)}$ by a total of j edge pairs split up between the two ends of the augmenting path given by $x^{(t)}$. We may change $x^{(t)}$ further but the above change is necessary to reduce $\ell(S^{(t)})$ by $2j$ edges. There are $j+1$ ways to split up the j edge pairs between the two ends of the augmenting path given by $x^{(t)}$. Hence, independent of $S^{(t)}, \ldots, S^{(0)}$,

$$\Pr\left[\ell(S^{(t+1)}) - \ell(S^{(t)}) \le -j \,\middle|\, S^{(t)}, \ldots, S^{(0)}\right] \le \frac{j+1}{m^{2j}}\,. \tag{5.5.4}$$

Consider the sequence $\{X^{(s)}\}_{s\in\mathbb{N}}$ with $X^{(s)} \in \mathbb{Z} \cup \{-\infty\}$ which is a pessimistic view on the sequence $\{\ell(S^{(t)})\}_{t\in\mathbb{N}}$ in the sense that $\{X^{(s)}\}_{s\in\mathbb{N}}$ is at most as likely as $\{\ell(S^{(t)})\}_{t\in\mathbb{N}}$ to reach k before 0.

Informally speaking, we construct $\{X^{(s)}\}_{s\in\mathbb{N}}$ from $\{\ell(S^{(t)})\}_{t\in\mathbb{N}}$ by (i) ignoring time steps where $\ell(S^{(t)})$ does not change at all; (ii) increasing $X^{(s)}$ by one whenever $\ell(S^{(t)})$ increases by at least one; (iii) decreasing $X^{(s)}$ by one whenever $\ell(S^{(t)})$ decreases by exactly one; and (iv) setting $X^{(s)}$ to $-\infty$ whenever $\ell(S^{(t)})$ decreases by at least two.

Formally, we define $X^{(s)}$ recursively. Let $t(0) := 0$ and $x^{(0)} := 1$. For $s \ge 1$, let $t(s)$ be the s-th point in time such that $\ell(S^{t(s)}) \ne \ell(S^{t(s)-1})$ and s_T the point in time such that $t(s_T) = T$. Then, for $s \in \mathbb{N}$ with $s < s_T$ let

$$X^{(s+1)} := \begin{cases} X^{(s)} + 1 & \text{if } \ell(S^{t(s)}) \ge \ell(S^{t(s)-1}) + 1\,, \\ X^{(s)} - 1 & \text{if } \ell(S^{t(s)}) = \ell(S^{t(s)-1}) - 1\,, \\ -\infty & \text{if } \ell(S^{t(s)}) \le \ell(S^{t(s)-1}) - 2\,. \end{cases}$$

We can infer from (5.5.2), (5.5.3), and (5.5.4) by increasing the bound in (5.5.4) to $5/m^4$ that for all $s < s_T$,

$$X^{(s+1)} := \begin{cases} X^{(s)} + 1 & \text{with probability at least } \frac{3}{5}(1 - \frac{1}{m^2+1})\,, \\ X^{(s)} - 1 & \text{with probability at most } \frac{2}{5}(1 - \frac{1}{m^2+1})\,, \\ -\infty & \text{with probability at most } \frac{1}{m^2+1}\,. \end{cases}$$

For $s \ge s_T$, we let the random variables $\{X^{(s)}\}_{s \ge s_T+1}$ be mutually independent such that

$$X^{(s+1)} := \begin{cases} X^{(s)} + 1 & \text{with probability } \frac{3}{5}(1 - \frac{1}{m^2+1})\,, \\ X^{(s)} - 1 & \text{with probability } \frac{2}{5}(1 - \frac{1}{m^2+1})\,, \\ -\infty & \text{with probability } \frac{1}{m^2+1}\,. \end{cases}$$

Let R be the random variable that denotes the first point in time $s \in \mathbb{N}$ such that $X^{(s)} \le 0$ or $X^{(s)} = k$. By definition of $X^{(s)}$, it holds that $\ell(S^{t(s)}) \ge X^{(s)}$ for all $s \le s_T$. Thus,

$$\Pr\left[\ell(S^{(T)}) = k\right] \ge \Pr\left[X^{(R)} = k\right].$$

In the remainder of this proof we show that $\Pr\left[X^{(R)} = k\right] \ge \delta$ for some constant $\delta > 0$.

By Markov's inequality (Theorem 4.1), it holds that $X^{(s)} > -\infty$ for all $s \le \frac{m^2+1}{2}$ with probability at least $1/2$. Suppose that indeed $X^{(s)} > -\infty$ for all $s \le \frac{m^2+1}{2}$. Then, until time $\frac{m^2+1}{2}$, the $X^{(s)}$'s perform a random walk on $\mathbb{Z}$ such that $X^{(s)}$ increases with probability at least 3/5 and decreases with probability at most 2/5.

For $i \in \{1, \dots, k-1\}$, let $b_i := \binom{i}{2}$ and let A_i be the event that $X^{(b_i)} \geq i$ and $X^{(s)} > 0$ for all $s \leq b_i$.

Let $i \in \{1, \dots, k-1\}$. We bound the probability of A_{i+1} conditioned on A_i. Since $b_{i+1} - b_i = \binom{i+1}{2} - \binom{i}{2} = i$, the event $X^{(b_{i+1})} \geq i+1$ conditioned on A_i implies that $X^{(b_i+1)}, \dots, X^{(b_{i+1})}$ are all positive.

Let $\mu := \mathrm{E}[X^{(b_{i+1})} \mid A_i]$ be the expected value of $X^{(b_{i+1})}$ conditioned on $X^{(b_i)} \geq i$ and $X^{(s)} > 0$ for all $s \leq b_i$. By the linearity of expectation,

$$\mu \geq i - \tfrac{2}{5}\, i + \tfrac{3}{5}\, i = \tfrac{6}{5}\, i\,.$$

By Theorem 4.13, we can apply the Chernoff bounds (Theorem 4.2) to probability that $X^{(b_{i+1})} \leq i$. Therefore,

$$\Pr\left[X^{(b_{i+1})} \leq i \,\middle|\, A_i\right] = \Pr\left[X^{(b_{i+1})} \leq \left(1 - \tfrac{1}{6}\right)\mu \,\middle|\, A_i\right] \leq \mathrm{e}^{-i/60}\,.$$

Thus, $\Pr[A_{i+1} \mid A_i] \geq 1 - \mathrm{e}^{-i/60}$ and

$$\Pr[A_k] = \Pr[A_0] \cdot \prod_{i=1}^{k-1} \Pr[A_{i+1} \mid A_i] \geq \prod_{i=1}^{k-1} \left(1 - \mathrm{e}^{-i/60}\right).$$

Let $L \in \mathbb{N}$ be minimal such that $\mathrm{e}^{-L/60} \leq 1/2$. Then, since $1 - a \geq e^{-2a}$ holds for all $a < 1/2$,

$$\begin{aligned}
\Pr[A_k] &\geq \left(1 - \mathrm{e}^{-1/60}\right)^L \cdot \prod_{i \in \mathbb{N}} \mathrm{e}^{-2\mathrm{e}^{-i/60}} \\
&= \left(1 - \mathrm{e}^{-1/60}\right)^L \cdot \mathrm{e}^{-2\sum_{i\in\mathbb{N}} \mathrm{e}^{-i/60}} \\
&= \left(1 - \mathrm{e}^{-1/60}\right)^L \cdot \mathrm{e}^{-\frac{2}{1-\mathrm{e}^{-1/60}}}\,.
\end{aligned}$$

Thus, $\Pr[A_k] > 2\delta$ where $\delta > 0$ is a constant. Since $b_k = \binom{k}{2} \leq \frac{m^2+1}{2}$, this implies that $\Pr[X^{(R)} = k] > \delta$ which concludes the proof. □

Using the previous two propositions, we finally show Theorem 5.19.

Proof of Theorem 5.19. We want to give a lower bound on $\mathrm{E}[T^*]$, the expected optimization time of the (1+1) $\mathrm{EA}_{\mathrm{MM}}$. By Proposition 5.20, Proposition 5.21, and Proposition 5.22, we know that there exists a constant $\delta > 0$ such that

$$\begin{aligned}
\mathrm{E}[T^* \mid x^{(0)} \neq M^*] &\geq \frac{1}{2m^3}\, \mathrm{E}[T^* \mid x^{(0)} = M^\circ] \\
&\geq \frac{1}{2m^3}\, \mathrm{E}[T_0 \mid \ell(S^{(0)}) = 1] \\
&\geq \frac{\delta}{2m^3}\, \mathrm{E}[T_0 \mid \ell(S^{(0)}) = k-1]\,.
\end{aligned}$$

We have seen in the proof of Proposition 5.22, that for all $1 \leq i \leq k-2$,

$$\mathrm{E}\big[\ell(S^{(t+1)}) - \ell(S^{(t)}) \mid \ell(S^{(t)}) = i\big] \geq \frac{3}{m^2} - \frac{2}{m^2} - \frac{k-1}{m^4} \geq \frac{1}{2m^2},$$

Furthermore, by (5.5.4) we have for all $1 \leq i \leq k-1$ and $j \geq 1$,

$$\Pr\left[\ell(S^{(t+1)}) - \ell(S^{(t)}) = -j \mid \ell(S^{(t)}) = i\right] \leq \frac{j+1}{m^{2j}} \leq \frac{1}{3^j}.$$

We apply Theorem 4.9 with $n = k-1$, $\varepsilon := 1/(2m^2)$, $\delta := 2$ and $r := 0$. Then, there exists two constants $\alpha, \beta > 0$ such that

$$\Pr[T_0 \leq \mathrm{e}^{\alpha(k-1)} \mid \ell(S^{(0)}) = k-1] \leq \mathrm{e}^{-\beta(k-1)}$$

and Theorem 5.19 follows by the law of total expectation. □

5.6. Shortest Path Problems

Another type of problems for which evolutionary algorithms have been extensively studied are *shortest path problems*. In particular, much attention has been paid to the *single-source shortest path (SSSP) problem* (Scharnow, Tinnefeld, and Wegener (2004); Doerr, Happ, and Klein (2007a); Baswana, Biswas, Doerr, Friedrich, Kurur, and Neumann (2009)) and the *all-pair shortest path (APSP) problem*, (Doerr, Happ, and Klein (2008a); Doerr and Theile (2009); Horoba and Sudholt (2009)).

Problem 5.23 (The Single-Source Shortest Path Problem).
Let $G = (V, E)$ be a strongly connected directed graph, let s be a distinguished source vertex in V, and let $w \colon E \to \mathbb{R}^+$ be a weight function. For $v \in V$, an optimal s-v-path in G is a directed path $P \subseteq E$ from s to v that minimizes $w(P) := \sum_{e \in P} w(e)$. The single-source shortest path problem asks for an optimal s-v-path in G for every $v \in V$.

A *directed spanning tree* $F \subseteq E$ with root s is a tree in G spanning all vertices of G such that the edges are directed away from s. It is well-known that there exist a set of optimal s-v-paths that forms a directed spanning tree with root s. Such a tree is called a *shortest path tree* of s (see, e. g., Mehlhorn and Sanders (2009); Cormen et al. (2001)).

From now on, we interpret the SSSP problem as the problem of finding a shortest path tree. That is, the search space $\mathcal{F}$ of the basic combinatorial (1+1) EA for the SSSP problem is the space of all directed spanning trees $F \subseteq E$ of G with root s.

We study two ways to define the order relation on $\mathcal{F}$ required by the basic combinatorial (1+1) EA. In this section, we analyze the single-criterion objective function that sums the weights of all paths. In Section 5.7, we analyze the multi-criteria optimization problem that optimizes the different path weights independently.

Let $F \in \mathcal{F}$. For all vertices $v \in V$, let P_v be the directed s-v-path in F and let $w(v, F) := w(P_v)$ be its weight (with $P_s := (s)$ and $w(s, F) := 0$). Then, the single-criterion objective function $f \colon 2^E \to \mathbb{R}_0^+$ for the SSSP problem on G is defined by $f(F) := \sum_{v \in V} w(v, F)$ for all directed spanning trees $F \subseteq E$ with root s.

Algorithm 5.24 ((1+1) EA$_{\text{SC-SSSP}}$). *The basic combinatorial (1+1) EA for the single-criterion single-source shortest path problem, the (1+1) EA$_{\text{SC-SSSP}}$, is an instance of Algorithm 5.2. Let $G = (V, E)$ be a strongly connected directed graph, s a source vertex of G, and $w \colon E \to \mathbb{R}^+$ a weight function on E. Then, the space of feasible search points $\mathcal{F}$ consists of all directed spanning trees of G with root s and $F \succeq H$ holds for $F, H \in \mathcal{F}$ if $f(F) \le f(H)$ holds.*

Note that $f(F)$ is not the same weight function as for the MST problem. In particular, $f(F)$ strongly depends on the choice of s. Thus, the analysis of the (1+1) EA$_{\text{SC-SSSP}}$ is not covered by that of the (1+1) EA$_{\text{MWB}}$ in Section 5.3. Again, we determine the drift in the objective function in each iteration of the algorithm (confer Baswana et al. (2009)).

Proposition 5.25. *Let $G = (V, E)$ be a strongly connected directed graph, s be a distinguished source vertex in V, and $w \colon E \to \mathbb{R}^+$ a weight function on E. Furthermore, let f_{opt} be the single-criterion objective value of a shortest path tree and let $\{x^{(t)}\}_{t\in\mathbb{N}}$ be the sequence of search points generated by the (1+1) EA$_{\text{SC-SSSP}}$ on (G, s, w). Then, regardless of $x^{(0)}$,*

$$\mathrm{E}[f(x^{(t)}) - f(x^{(t+1)}) \mid f(x^{(t)})] \ge \frac{f(x^{(t)}) - f_{\text{opt}}}{\mathrm{e}\,|E|^2\,|V|}\,.$$

Proof. Let $p := (1 - 1/|E|)^{|E|-2}\,|E|^{-2}$ and let S be a shortest path tree, that is, $f(S) = f_{\text{opt}}$. Furthermore, let $F := x^{(t)}$ with $t \in \mathbb{N}$. If $f(F) = f_{\text{opt}}$ the statement follows trivially. Thus, suppose $f(F) > f(S)$. Then,

$$f(F) - f(S) = \sum_{v\in V} \big(w(v, F) - w(v, S)\big)\,.$$

Thus, there exists a $v \in V$ such that

$$w(v, F) - w(v, S) \ge \frac{f(F) - f(S)}{|V|}\,.$$

Let $P := (u_0, e_1, u_1, \ldots, u_{k-1}, e_k, u_k)$ with $k \in \mathbb{N}$ be the s-v-path in S with $s = u_0$ and $v = u_k$. To each edge e_j of P, we assign a contribution $c(e_j)$ to $w(v, F) - w(v, S)$ by defining

$$c(e_j) := \big(w(u_j, F) - w(u_j, S)\big) - \big(w(u_{j-1}, F) - w(u_{j-1}, S)\big)$$

for all $j \in \{1, \ldots, k\}$. This implies

$$\sum_{j=1}^{k} c(e_j) = w(v, F) - w(v, S)\,.$$

Note that since $e_j \in P$, it holds for all $j \in \{1, \ldots, k\}$ that

$$c(e_j) = w(u_j, F) - w(u_{j-1}, F) - w(e_j)\,. \tag{5.6.1}$$

Among the edges of P, we are particularly interested in those with strictly positive contribution. We call their index set J, that is,

$$J := \{j \in \{1, \ldots, k\} \mid c(e_j) > 0\} .$$

Let $j \in J$. Then it follows from Equation 5.6.1 that $e_j \notin F$ and u_{j-1} is not in the directed subtree of F which has u_j as its root. Let $\tilde{e}_j := (\tilde{u}_{j-1}, u_j)$ be the last edge of the s-u_j-path in F. Let $F_j = F \cup \{e_j\} \setminus \{\tilde{e}_j\}$ be the directed spanning tree with root s that is obtained by detaching the subtree of F rooted at u_j from $\tilde{u}_{j-1}$ and re-attaching it at u_{j-1}.

Then, $w(u_j, F_j) = w(u_{j-1}, F) + w(e_j)$ and thus again by Equation 5.6.1,

$$c(e_j) = w(u_j, F) - w(u_j, F_j) .$$

Since $w(u, F)$ is at least $w(u, F_j)$ for all $u \in V$, we obtain

$$f(F) - f(F_j) \geq c(e) .$$

Now the remainder of the proof is analogue to that of Proposition 5.9. For $j \in J$ let

$$I_j := \begin{cases} 1 & \text{if } x^{(t+1)} = F_j , \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Pr[I_j = 1] = p$,

$$f(x^{(t+1)}) - f(x^{(t)}) \geq \sum_{j \in J} c(e_j)\, I_j ,$$

and

$$\mathrm{E}[f(x^{(t+1)}) - f(x^{(t)}) \mid f(x^{(t)})] \geq \sum_{j \in J} c(e_j)\, p \geq \frac{f(F) - f(S)}{\mathrm{e}\, |E|^2\, |V|} .$$

□

This result indicates that $g\colon \mathcal{F} \to \mathbb{R}_0^+$ defined by $g(F) = f(F) - f_{\text{opt}}$ for all $f \in \mathcal{F}$ is a suitable potential function. Using Theorem 4.5, we bound the expected optimization time of the (1+1) $\text{EA}_{\text{SC-SSSP}}$.

Theorem 5.26 (confer Baswana et al. (2009)). *Let $G = (V, E)$ be a strongly connected directed graph, s be a distinguished source vertex in V, and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Let f_{opt} be the single-criterion objective value of a shortest path tree and $f_{2nd\text{-opt}}$ be the minimal single-criterion objective value over all directed spanning trees with root s that do not have objective value f_{opt}.*

Furthermore, let T be the optimization time of the (1+1) $\text{EA}_{\text{SC-SSSP}}$ on (G, s, w). Then,

$$\mathrm{E}[T \mid x^{(0)}] \leq \mathrm{e}\, |E|^2\, |V| \left(1 + \ln \frac{f(x^{(0)}) - f_{\text{opt}}}{f_{2nd\text{-opt}} - f_{\text{opt}}}\right) .$$

For integer weights polynomially bounded in $|E|$, the previous theorem shows that the expected optimization time of the (1+1) $\mathrm{EA}_{\text{SC-SSSP}}$ is $\mathcal{O}(|E|^2\,|V|\,\ln|V|)$. Note that for general weights we cannot apply Theorem 5.11 to give a weight-independent upper bound.

We will see in Section 5.9 that by using a more sophisticated problem representation this bound can be decreased to $\mathcal{O}(|E|\,|V|\,\ln|V|)$. In comparison, the runtime of the problem-specific algorithm by Dijkstra (1959) is $\mathcal{O}(|E| + |V|\ln|V|)$.

As for the minimum spanning tree and maximum weight basis problems, it is an open question whether the asymptotic optimization time depends on the choice of the weights and of the graph for non-trivial instances. In this context, it would be interesting to see whether it is possible to adapt Theorem 5.11 to the case of the SSSP problem.

5.7. Multi-Criteria Optimization

In Section 5.7, we analyzed the how basic combinatorial (1+1) EA optimizes the SSSP problem using a single-criterion objective function. Let us briefly recall the setting in Section 5.6. Given a graph $G = (V, E)$ and a source $s \in V$, the search space $\mathcal{F}$ consists of all directed spanning trees of G with root s. For a tree $F \in \mathcal{F}$ and a vertex $v \in V$, we denote by $w(v, F)$ the sum of the weights of the edges of the unique s-v-path in F.

In Section 5.7, we considered the single-criterion objective function $f\colon \mathcal{F} \to \mathbb{R}_0^+$ with $f(F) := \sum_{v\in V} w(v, F)$. Now, we study the multi-criterion approach that optimizes each $w(v, \cdot)$ independently.

Algorithm 5.27 ((1+1) $\mathrm{EA}_{\text{MC-SSSP}}$). *The basic combinatorial (1+1) EA for the multi-criteria single-source shortest path problem, the (1+1) $\mathrm{EA}_{\text{MC-SSSP}}$, is an instance of Algorithm 5.2. Let $G = (V, E)$ be a strongly connected directed graph, s a source vertex of G, and $w\colon E \to \mathbb{R}^+$ a weight function on E. Then, the space of feasible search points $\mathcal{F}$ consists of all directed spanning trees of G with root s and $F \succeq H$ holds for $F, H \in \mathcal{F}$ if $w(v, F) \le w(v, H)$ for all $v \in V$ holds.*

The multi-criteria setting strongly simplifies the analysis. Once the (1+1) EA finds an optimal s-v-path, the objective functions ensures that it is never replaced by a sub-optimal path (which in general is not true in the single-criterion setting).

In the following theorem, the optimization time depends on the *unweighted edge radius* $\ell_G(s)$. This is the minimal value $\ell \in \mathbb{N}$ such that for every vertex $v \in V$ there exists an optimal s-v-path on at most ℓ edges. That is,

$$\ell_G(s) := \max_{v\in V}\left(\min_{P\colon P \text{ is optimal } s\text{-}v\text{-path}} \big(|E(P)|\big)\right).$$

Theorem 5.28 (confer Doerr et al. (2007a)). *Let $G = (V, E)$ be a strongly connected directed graph, s be a distinguished source vertex in V, and let $w\colon E \to \mathbb{R}^+$ be a weight function on E.*

Furthermore, let T be the optimization time of the (1+1) $\mathrm{EA}_{\text{MC-SSSP}}$ (Algorithm 5.27) on (G, s, w). Then for all $\varepsilon > 0$,

$$\Pr\left[T \le (1+\varepsilon)(2+\sqrt{3})\,\mathrm{e}\,|E|^2\,\max\{\ell_G(s)\,,\,\ln|V|\}\,\middle|\,x^{(0)}\right] \ge 1 - |V|^{-\varepsilon}.$$

Proof. Let $\varepsilon > 0$ and $x^{(0)}$ be fixed. Let $p := (1 - 1/|E|)^{|E|-2}\,|E|^{-2}$, let $n := |V|$, and let $\ell^* := \max\{\ell_G(s)\,,\, \ln n\}$. For all $v \in V$, let $w_{\mathrm{opt}}(v)$ be the weight of an optimal s-v-path and let T_v be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which the s-v-path in $x^{(t)}$ is optimal. We first show that

$$\Pr[T_v > (1+\varepsilon)\,(2+\sqrt{3})\,p^{-1}\,\ell^*] \le n^{-(1+\varepsilon)}$$

holds for all $v \in V$. Let $v \in V$ and let $P := (u_0, e_1, u_1, \ldots, u_{k-1}, e_k, u_k)$ be an optimal s-v-path on $k \le \ell_G(s)$ edges with $s = u_0$ and $v = u_k$.

Consider a point in time $t \le T_v$ and let $j(t) \in \{1, \ldots, k\}$ be minimal such that $w(u_{j(t)}, x^{(t-1)}) \ne w_{\mathrm{opt}}(u_{j(t)})$. Like in the single-criterion case, we know that the edge $e_{j(t)} := (u_{j(t)-1}, u_{j(t)})$ is not in $x^{(t-1)}$. Moreover, there has to exist another edge $\tilde{e}_{j(t)} = (\tilde{u}_{j(t)-1}, u_{j(t)})$ with $x^{(t-1)} \cup \{e_{j(t)}\} \setminus \{\tilde{e}_{j(t)}\} \in \mathcal{F}$ (see proof of Proposition 5.25). We say $x^{(t)}$ is a *pessimistic improvement* if $x^{(t)} = x^{(t-1)} \cup \{e_{j(t)}\} \setminus \{\tilde{e}_{j(t)}\}$.

If $x^{(t)}$ is a pessimistic improvement, then $w(u_i, x^{(t)}) = w_{\mathrm{opt}}(u_i)$ for all $i < j(t)$. Let $j \in \{1, \ldots, k\}$. Then there exists at most one point in time $t \in \mathbb{N}$ for which $j(t) = j$ and $x^{(t+1)}$ is a pessimistic improvement.

Thus, if k pessimistic improvements occur then $w(v, x^{(t)}) = w_{\mathrm{opt}}(v)$ holds with certainty. Note that in general the converse is not true. At no time $t \in \mathbb{N}$, the s-u_i-sub-path of P can be guaranteed to be a subgraph of $x^{(t)}$ for $i > 1$.

Now we make use of the fact that for $t \le T_v$, the events that the $x^{(t)}$'s are pessimistic improvements are mutually independent (although the positions $j(t)$ of these improvements may be highly dependent). More precisely, the probability that $x^{(t)}$ and $x^{(t-1)}$ differ only in the edges $e_{j(t)}$ and $\tilde{e}_{j(t)}$ is exactly p and this holds for all outcomes of $x^{(0)}, \ldots, x^{(t-1)}, x^{(t+1)}, \ldots, x^{(T_v)}$.

For $t \le T_v$, let $I^{(t)} \in \{0, 1\}$ be the indicator variable that is one if $x^{(t)}$ is a pessimistic improvement. Then $\Pr[I^{(t)} = 1] = p$ and the $I^{(t)}$ are mutually independent for $t \le T_v$. Now, for $t > T_v$, we let $I^{(t)}$ be auxiliary random variables in $\{0, 1\}$ which are mutually independent and also assume the value one with probability p.

Recall the observation that if there have been k pessimistic improvements at time $r := (1+\varepsilon)\,(2+\sqrt{3})\,p^{-1}\,\ell^*$ then $w(v, x^{(r)}) = w_{\mathrm{opt}}(v)$. Let $Z := \sum_{t=1}^{r} I^{(t)}$. Then, the event "$Z \ge k$" implies the event "$T_v \le r$". In terms of probabilities,

$$\Pr[Z \ge k] \le \Pr[T_v \le r]\,.$$

Thus, since $k \le \ell_G(s) \le \ell^*$,

$$\Pr[T_v > r] \le \Pr[Z < k] \le \Pr[Z < \ell^*]\,.$$

Now, $\mathrm{E}[Z] = r\,p = (1+\varepsilon)\,(2+\sqrt{3})\,\ell^*$. Hence,

$$\Pr[T_v > r] \le \Pr\left[Z < (1+\varepsilon)^{-1}\,(2+\sqrt{3})^{-1}\,\mathrm{E}[Z]\right].$$

We can now apply the Chernoff bounds (Theorem 4.2) to the right hand-side of the previous inequality. Since $\ell^* \ge \ln n$, we obtain

$$\Pr[T_v > r] \le \mathrm{e}^{-(1+\varepsilon)\,\ell^*} \le n^{-(1+\varepsilon)}\,.$$

Finally, since the previous inequality holds for all $v \in V$, the statement follows by the union bound argument

$$\Pr[T > r] = \Pr\Big[\bigcup_{v \in V} (T_v > r)\Big] \leq n^{-\varepsilon}.$$

□

5.8. Permutation Based Search Spaces

All search spaces we investigated so far were composed of subsets of a given base set E, mostly the edge set of a graph. Because of this, we were able to apply the basic combinatorial (1+1) EA. However, sometimes other representations of the search space are more natural and, as we will see, can prove to be more efficient.

In this section, we introduce a generic (1+1) EA for general search spaces. On the example of the classical *Euler tour problem*, we discuss different ways of how this algorithm can be applied to the search space of permutations.

Problem 5.29 (The Euler Tour Problem).
Let $G = (V, E)$ be an Eulerian graph (G is connected and every vertex of G has even degree). The Euler tour problem asks for a closed walk in G that uses each edge exactly once.

We represent walks in G by permutations of the edge set E (confer Scharnow et al. (2004)). Let $E = \{e_1, \dots, e_m\}$ with $m = |E|$ and let $\mathcal{S}_m$ be the space of all permutations of the numbers $1, \dots, m$. Furthermore, let $W := (u_0, f_1, \dots, f_k, u_k)$ be a walk in G of length $k \in \mathbb{N}$.

We identify W with all permutations $\sigma \in \mathcal{S}_m$ such that $u_k \notin e_{\sigma(k+1)}$ and such that $f_j = e_{\sigma(j)}$ holds for all $j \in \{1, \dots, k\}$ (permutations with $u_k \in e_{\sigma(k+1)}$ correspond to longer walks). For a permutation σ, let $k = k(\sigma)$ be the maximal integer such that $W_\sigma := (u_0, e_{\sigma(1)}, \dots, e_{\sigma(k)}, u_k)$ is a proper walk. Then W_σ is the walk that corresponds to σ.

Euler's Theorem (Euler (1741)) guarantees that in an Eulerian graph a walk of length m exists and moreover that all such walks are closed. Thus, we may rephrase the Euler tour problem as the problem to optimize the objective function $k \colon \mathcal{S}_m \to \mathbb{N}$ as defined above over the search space $\mathcal{S}_m$ (where k depends on the structure of G).

We now formulate the *generic* (1+1) EA which works on arbitrary search spaces. For this, we define how to randomly generate the candidate search point $y^{(t+1)}$ from the current search point $x^{(t)}$. We call the process *variation*.

Definition 5.30 (variation operator ϕ). *Let $\mathcal{S}$ be a finite search space. A variation operator is a sampling procedure that generates a random search point in $\mathcal{S}$ according to a distribution that is based on a given search point in $\mathcal{S}$.*

The generic (1+1) EA mimics the two defining features of the basic combinatorial (1+1) EA. These are that (i) the candidate search point $y^{(t+1)}$ is likely to be chosen in a local neighborhood of $x^{(t)}$ and (ii) each point in the search space can be chosen as $y^{(t+1)}$ with positive probability. The first properties ensures that most of the time the algorithm explores the search space locally. The second property guarantees that the algorithm eventually finds a global optimum.

Let us recall how variation is performed by the basic combinatorial (1+1) EA. Each element $e \in E$ indicates a location for a potential variation by adding or removing e from $x^{(t)}$. The candidate solution $y^{(t+1)}$ is generated by performing a variation of $x^{(t)}$ at location e with probability $1/|E|$ independently for all edges $e \in E$.

For the search space of permutations $\mathcal{S}_m$, several notions of locality are possible. The canonical local variation operator is the *exchange* variation operator ϕ_{exchange}. Let $\sigma \in \mathcal{S}_m$ be a permutation. Then $\phi_{\text{exchange}}(\sigma)$ is generated by choosing two positions a, b uniformly at random from $\{1, \ldots, m\}$ and transposing them in σ.

A second local variation operator is the *jump* variation operator ϕ_{jump}. Again, let $\sigma \in \mathcal{S}_m$ be a permutation. This time, ϕ_{jump} is generated by choosing two positions a, b uniformly at random from $\{1, \ldots, m\}$, removing the element at position a from σ, and reinserting it so that it becomes position b.

In both cases, the respective variation operator is local in the informal sense of property (i). However, unlike the case of the basic combinatorial (1+1) EA, two different exchanges might change the same position of the permutation. Thus, we cannot perform several local variations simultaneously.

To obtain property (ii), we first choose a random number k according to Pois(1), the Poisson distribution with parameter 1. Then, we perform k sequential random local variations. By this, we transform a local variation operator into a variation operators that also satisfies property (ii). We specifically choose the Poisson distribution to simulate the basic combinatorial (1+1) EA. There, the number of local variations (changed elements) is governed by the Binomial distribution which is known to converge to the Poisson distribution[3].

Algorithm 5.31 (The Generic (1+1) EA).
Let $\mathcal{S}$ be a finite search space. Furthermore, let ϕ be a local variation operator on $\mathcal{S}$, let $\succeq$ be a partial order relation on $\mathcal{S}$, and let $x^{(0)} \in \mathcal{S}$ be an initial search point. Given $(\mathcal{S}, \phi, \succeq, x^{(0)})$, the generic (1+1) EA iteratively generates a sequence of search points $\{x^{(t)}\}_{t \in \mathbb{N}}$ in $\mathcal{S}$ by the following procedure.

For all $t \in \mathbb{N}$ with $t \geq 1$, the generic (1+1) EA generates a random number k according to Pois(1). *Next, the generic (1+1) EA generates the candidate search point $y^{(t)} := \phi^k(x^{(t-1)})$ by k successive applications of ϕ to $x^{(t-1)}$. Afterwards,*

$$x^{(t)} := \begin{cases} y^{(t)} & \textit{if } y^{(t)} \succeq x^{(t-1)}\,, \\ x^{(t-1)} & \textit{otherwise.} \end{cases}$$

The *optimization time* T of the generic (1+1) EA for $(\mathcal{S}, \phi, \succeq, x^{(0)})$ is the random variable that describes the first point in time $t \in \mathbb{N}$ for which $x^{(t)}$ is optimal.

Neumann (2008) showed that the generic (1+1) EA for the Euler tour problem

[3] In the results referenced in this chapter, the number of successive variation operators applied in each single variation step is often distributed according to $1 + \text{Pois}(1)$ instead of Pois(1). From the analytic point of view, this choice is artificial as it deviates from the original purpose of simulating the basic combinatorial (1+1) EA. Furthermore, this difference does not have an effect on the order of magnitude of the expected runtimes considered in this chapter.

using the exchange variation operator ϕ_{exchange} has at least exponential expected optimization time.

Theorem 5.32 (Neumann (2008)). *Let $n \geq 3$. Then, there exists an Eulerian graph $G = (V, E)$ of size n such that the expected optimization time of the generic (1+1) EA for the Euler tour problem on G using the local exchange variation operator ϕ_{exchange} with $x^{(0)}$ chosen uniformly at random is $2^{\Omega(|V|)}$.*

However, the *jump* variation operator ϕ_{jump} results in a polynomially bounded expected optimization time of the generic (1+1) EA for the Euler tour problem.

Theorem 5.33 (Neumann (2008)). *Let $G = (V, E)$ be an Eulerian graph. Then the expected optimization time of the generic (1+1) EA for the Euler tour problem on G using the local exchange variation operator ϕ_{jump} is $\mathcal{O}(|E|^5)$ independent of $x^{(0)}$. For all $n \geq 3$, there exists an Eulerian graph $G = (V, E)$ on n vertices such that the expected optimization time of the generic (1+1) EA for the Euler tour problem on G using the local exchange variation operator ϕ_{jump} with $x^{(0)}$ chosen uniformly at random is $\Omega(|E|^5)$.*

In Doerr, Hebbinghaus, and Neumann (2007b) it is shown that if $a = 1$ is fixed in ϕ_{jump}, then the expected optimization time of the generic (1+1) EA for the Euler tour problem drops to $\Theta(|E|^3)$.

From these three examples, we conclude that for non-Boolean search spaces (like the space of permutations) the choice of locality can severely influence the optimization time.

However, an expected optimization time of $\Theta(|E|^3)$ for the Euler tour problem is still far off from the linear runtime of the problem-specific algorithm by Hierholzer (1873). In the following Section, we will see how it is possible to further reduce the expected optimization time of the generic (1+1) EA to $\mathcal{O}(|E| \ln |E|)$.

5.9. Asymmetric and Adjacency-Based Variation Operators

In Section 5.8, we have seen that the choice of the local variation operator can have significant influence on the optimization time of the generic (1+1) EA. In this section we revisit the problems of finding a maximum weight basis, a single-source shortest path tree, an Euler tour, or an all-pair shortest path set. We study how alternative representations and variation operators for these problems influence the optimization time.

Consider the basic (1+1) EA for the maximum weight basis problem and recall the proof of Proposition 5.9. There, the bound on the drift is dominated by the probability $p_f = (1 - m^{-1})^{m-2} m^{-2}$ (where $m = |E|$) to exchange a particular element in a basis of E by a particular element that is not in E. However, we know that all bases are of the same size, namely the rank $r = r(M)$ of the matroid (which is $|V| - 1$ in the case of spanning trees).

We modify the distribution of the variation operator in the basic combinatorial (1+1) EA (confer Jansen and Sudholt (2010); Reichel and Skutella (2007)). The *asymmetric variation operator* is defined as follows. Instead of choosing each element e

in $x^{(t-1)} \triangle y^{(t)}$ independently with probability $1/m$, we independently choose e with probability $1/r$ if $e \in x^{(t-1)}$ with probability $1/(m-r)$ if $e \notin x^{(t-1)}$. This way, in expectation $x^{(t-1)} \triangle y^{(t)}$ contains exactly two elements, one in $x^{(t)}$ and in $E \setminus x^{(t)}$. Consequently, for a particular pair of such elements

$$p_f = \Big(1 - \frac{1}{r}\Big)^{r-1} \Big(1 - \frac{1}{m-r}\Big)^{m-r-1} \frac{1}{r(m-r)} \geq \frac{1}{\mathrm{e}^2\, r\, m} .$$

Thus, following the proof of Proposition 5.9 the bound of expected optimization time in Theorem 5.10 drops by a factor of m/r.

Theorem 5.34 (confer Reichel and Skutella (2007)). *Let $M = (E, \mathcal{F})$ be a matroid and let $w \colon E \to \mathbb{R}^+$ be a weight function on E. Let w_{opt} be the weight of a maximum weight basis and $w_{\text{2nd-opt}}$ be the maximum weight over all bases that do not have weight w_{opt}. Furthermore, let T be the optimization time of the (1+1) EA for the maximum weight basis problem on (M, w) using the asymmetric variation operator defined above. Then,*

$$\mathrm{E}[T \mid x^{(0)}] \leq \mathrm{e}^2\, r(M)\, |E| \left(1 + \ln \frac{w_{\text{opt}} - w(x^{(0)})}{w_{\text{opt}} - w_{2nd\text{-opt}}}\right)$$

Let us now turn to the SSSP problem. Recall the problem of finding a single-source shortest path tree according to a source vertex s in a strongly connected directed graph $G = (V, E)$ (Section 5.6). As search space, we considered all directed spanning trees with root s. We represented them as subsets of E, rejecting all subsets that do not from such a tree.

Let us take a closer look at the proof of Theorem 5.26. Consider the factor $\mathrm{e}\,|E|^2$ in the upper bound on the expected optimization time of the basic combinatorial (1+1) EA. We observe that it results from the probability p to change the position of a subtree in the current search point (see proof of Proposition 5.26).

In the representation by edge sets, this is done by removing the edge that attaches the subtree and inserting a new edge that re-attaches the subtree while leaving the other edges unchanged. Thus,

$$p = \Big(1 - \frac{1}{|E|}\Big)^{|E|-2} \frac{1}{|E|^2} \geq \frac{1}{\mathrm{e}\,|E|^2} .$$

Hence, the expected time for a particular subtree relocation is at most $\mathrm{e}\,|E|^2$.

To reduce the upper bound on the expected time to relocate a subtree, we choose a more suited representation of the search points. For this, we regard a data-structure commonly used to represent graphs — adjacency lists. The *adjacency list* L of a directed graph $G = (V, E)$ stores for every vertex $v \in V$ the sub-list $L(v) \subseteq E$ of all *incoming* directed edges $(w, v) \in E$.

A natural way to represent directed rooted spanning trees is to assign to each vertex the edge by which it is attached to its predecessor in the tree. Thus, in such a representation, we distinguish for each vertex $v \in V$ one edge (w, v) in its sub-list $L(v)$.

Let $\mathcal{L}$ be the space that consists of all sets of edges F such that there exists exactly one edge from $L(v)$ in F for every vertex $v \in V \setminus \{s\}$. Note that not every set in $\mathcal{L}$

corresponds to a directed spanning tree. In particular, a set in $\mathcal{L}$ might define a graph which contains directed cycles.

The search space $\mathcal{S}_{\text{SSSP}}$ consists of all sets in $\mathcal{L}$ that correspond to directed spanning trees rooted at s.

The local variation operator ϕ_{SSSP} on a set $F \in \mathcal{S}_{\text{SSSP}}$ generates the random set $F' \in \mathcal{S}_{\text{SSSP}}$ as follows. First, an edge e is chosen uniformly at random from $E \setminus L(s)$. Then, there exists a unique vertex $v \in V \setminus \{s\}$ such that $e \in L(v)$ and a unique edge $e' \in F \cap L(v)$. If $F \cup \{e\} \setminus \{e'\}$ represents a directed spanning tree rooted at s, then $F' = F \cup \{e\} \setminus \{e'\}$. Otherwise, $F' = F$.

We can now analyze the generic (1+1) EA for the SSSP problem on the search space $\mathcal{S}_{\text{SSSP}}$ and the local variation operator ϕ_{SSSP}. Let $\succeq_s$ and $\succeq_m$ be the partial orders on $\mathcal{S}_{\text{SSSP}}$ defined by the single-criterion objective function f in Section 5.6 and the multi-criteria partial order $\succeq$ in Section 5.7, respectively. Then the (1+1) EA using adjacency lists to optimize the single-criterion or multi-criteria SSSP problem is the generic (1+1) EA on the search space $\mathcal{S}_{\text{SSSP}}$ with variation operator ϕ_{SSSP} and an objective defined by the order relation $\succeq_s$ or $\succeq_m, x^{(0)}$, respectively.

It turns out that the proofs in Section 5.6 and Section 5.7 are still correct if we change the probability p to relocate a fixed subtree to a specific position. Using the local variation operator ϕ_{SSSP}, the lower bound on this probability increases by a factor of $|E|$ to

$$p' = \frac{1}{\mathrm{e}\,|E|}\,.$$

Thus, we get the following results which are the equivalents of Theorem 5.26 and Theorem 5.28 for the generic (1+1) EA using the local mutation operator ϕ_{SSSP}.

Theorem 5.35 (Doerr and Johannsen (2010)). *Let $G = (V, E)$ be a strongly connected directed graph, s be a distinguished source vertex in V, and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Let f_{opt} be the single-criterion objective value of a shortest path tree and $f_{2nd\text{-opt}}$ be the minimal single-criterion objective value over all directed spanning trees with root s that do not have objective value f_{opt}.*

Let $x^{(0)} \in \mathcal{S}_{SSSP}$ and let T_s and T_m be the optimization times of the (1+1) EA on the single-criterion (multi-criteria, respectively) SSSP problem using adjacency lists. Then, for all $\varepsilon > 0$,

$$\mathrm{E}[T_s \mid x^{(0)}] \le \mathrm{e}\,|E|\,|V| \left(1 + \ln \frac{f(x^{(0)}) - f_{\text{opt}}}{f_{2nd\text{-opt}} - f_{\text{opt}}}\right)$$

and

$$\Pr\left[T_m \le (1+\varepsilon)(2+\sqrt{3})\,\mathrm{e}\,|E|\,\max\{\ell_G(s)\,,\,\ln|V|\} \,\middle|\, x^{(0)}\right] \ge 1 - |V|^{-\varepsilon}\,.$$

Next, we consider the Euler tour problem. Also for this problem there exist a superior representation based on adjacency lists (Doerr, Klein, and Storch (2007c); Doerr and Johannsen (2007b)). As described in Section 5.8, this problem is defined on an undirected Eulerian graph $G = (V, E)$. The corresponding adjacency list L stores for every vertex $v \in V$ the sub-list $L(v)$ of edges incident with v. This way, the edge $\{v, w\}$ occurs twice, in $L(v)$ and in $L(w)$. Note, since G is Eulerian, all sub-lists are of even size.

Consider a walk $W = (u_0, f_1, \ldots, f_k, u_k)$ in G. Then W can be represented by all pairs of successive edges $\{f_1, f_2\}, \{f_2, f_3\}, \ldots, \{f_{k-1}, f_k\}$. For each pair $\{f_i, f_{i+1}\}$, there exists exactly one sub-list of L which contains the two edges f_i and f_{i+1}, namely $L(u_i)$. Thus, we identify the walk W with pairings of edges in the sub-lists of L.

The search space $\mathcal{S}_{\text{Euler}}$ consist of all complete pairings of edges in L. A *complete pairing* is a set of edge pairs such that (i) the edges of a pair are in the same sub-list of L and (ii) each edge in a sub-list of L belongs to exactly one pair. Such a pairing always exists since all sub-lists are of even lengths. Moreover, a random pairing is easy to generate: for each sub-list, we successively pair two random vertices and then remove them from the list.

Since each edge belongs to exactly two pairs, the pairs partition the edge sets into disjoint and cyclically ordered sets. We call these cyclically ordered sets *tours*. Note that if we distinguish some vertex of a tour as start-/end-vertex, then the tour becomes a closed walk.

We have just seen that a complete pairing corresponds to a disjoint partition of the edge set into tours. Let $k\colon \mathcal{S}_{\text{Euler}} \to \mathbb{N}$ be the function that counts the number of tours $k(x)$ in the partition corresponding to a search point $x \in \mathcal{S}_{\text{Euler}}$. Whenever $k(x) = 1$, the partition corresponding to x contains only one tour which has to be an Euler tour. Consequently, we reformulate the Euler tour problem over $\mathcal{S}_{\text{Euler}}$ as the problem of minimizing k over $\mathcal{S}_{\text{Euler}}$.

The local variation operator ϕ_{Euler} that generates a random search point $y \in \mathcal{S}_{\text{Euler}}$ based on the search point $x \in \mathcal{S}_{\text{Euler}}$ is defined as follows. We choose the two edges e and e' uniformly at random from all pairs of edges that are in the same sub-list of L. In x, each of these two edges is paired with a second edge, say $\{e, f\}$ and $\{e', f'\}$ are paired. Then y is generated by removing the pairs $\{e, f\}$ and $\{e', f'\}$ from x and adding the pairs $\{e, e'\}$ and $\{f, f'\}$ in return.

The local variation operator ϕ_{Euler} that generates a random search point $y \in \mathcal{S}_{\text{Euler}}$ based on the search point $x \in \mathcal{S}_{\text{Euler}}$ is defined as follows. We choose an edge e uniformly at random from E. Then, we choose with probability $1/2$ one of its end-vertices v and w (say v). Finally, we choose a second edge e' uniformly at random from $L(v)$. In x, each of these two edges is paired with a second edge, say $\{e, f\}$ and $\{e', f'\}$. Then y is generated by removing the pairs $\{e, f\}$ and $\{e', f'\}$ from x and adding the pairs $\{e, e'\}$ and $\{f, f'\}$ in return.

The (1+1) EA using adjacency lists to solve the Euler tour problem is defined to be the generic (1+1) EA on the search space $\mathcal{S}_{\text{Euler}}$ with the variation operator ϕ_{Euler} and an objective defined by the partial order relation $\succeq_{\text{Euler}}$ given by the function $k\colon \mathcal{S}_{\text{Euler}} \to \mathbb{N}$ above.

Theorem 5.36 (Doerr and Johannsen (2007b)). *Let $G = (V, E)$ be an Eulerian graph. Furthermore, let $x^{(0)} \in \mathcal{S}_{\text{Euler}}$ and T be the optimization time of the (1+1) EA using adjacency lists to solve the Euler tour problem. Then,*

$$\mathrm{E}[T \mid x^{(0)}] \le \mathrm{e}\,|E| \ln |E| \,.$$

Proof. Let $t < T$ and $x^{(t)}$ be the t-th search point of the (1+1) EA using adjacency lists to solve the Euler tour problem. Then $x^{(t)}$ corresponds to a partition of E into $g(x^{(t)})$ tours and $g(x^{(t)})$ is at least two.

Let τ be one of these tours. Since $t < T$ and G is connected, τ shares at least one vertex v with another tour in the partition. Suppose that ℓ is the size of the sub-list and k of edges in $L(v)$ belong to τ. Since there exist at least one pairing in $L(v)$ for each tour visiting v, we have that $\ell \geq 4$ and $2 \leq k \leq \ell - 2$.

The probability, that the (1+1) EA operator performs exactly one local variation and merges τ at v with a second tour by is

$$p = \frac{1}{\mathrm{e}} \frac{k}{|E|} \cdot \frac{\ell - k}{\ell} \geq \frac{1}{\mathrm{e}\,|E|}\,.$$

For each of the $g(x^{(t)})$ tours there exists at least one vertex v at which such a merging may occur. Thus, the probability to merge two tours is at least $\frac{g(x^{(t)})}{\mathrm{e}\,|E|}$. Note that we do not over-count since we choose ordered pairs of edges (e, e') for variation. Hence, in expectation the number of tours decreases by at least $\frac{g(x^{(t)})}{\mathrm{e}\,|E|}$. The statement follows by Theorem 4.5. □

5.10. Population and Recombination

In this section we study the *all-pair shortest path problem* (APSP). In the context of evolutionary algorithms, this problem has been studied by Doerr et al. (2008a) and Doerr and Theile (2009). The main insight of these works is that these algorithms perform better if they also apply recombination of two search points instead of variation of single search points only. Further theoretical work on the influence of recombination on the runtimes of evolutionary algorithms can be found in Fischer and Wegener (2004) and Watson and Jansen (2007).

Problem 5.37 (The All-Pair Shortest Path (APSP) Problem).
Let $G = (V, E)$ be a strongly connected directed graph and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. For all $v, w \in V$, an optimal v-w-path in G is a directed path $P \subseteq E$ from v to w minimizing $w(P) := \sum_{e \in P} w(e)$. The all-pair shortest path problem asks for an optimal v-w-path in G for every ordered pair (v, w) of distinct vertices $v, w \in V$.

A solution to the APSP problem is a set of optimal paths. Consequently, we apply an evolutionary algorithm that maintains in each iteration a set or multi-set of search points rather than a single search point. Moreover, for the APSP problem we observe that the concatenation of an optimal v-u-path with an optimal u-w-path often results in an optimal v-w-path.

In the context of evolutionary algorithms, we call the multi-set of current search points the *population* and the generation of a (random) search point from two others *recombination*. In our case, we perform recombination by the concatenation of two paths.

The concatenation $P \circ Q$ of two paths $P = (v, \ldots, w)$ and $Q = (w, \ldots, u)$ is the walk $(v, \ldots, w, \ldots, u)$. Note, that $P \circ Q$ may visit vertices more than once and thus need not be a path.

The following evolutionary algorithm generates a single candidate search point in each iteration. This is done by using either variation or recombination. In this, the chosen method depends on the recombination probability $p_{\mathrm{R}} \in [0,1]$. More precisely, the algorithm samples from the Bernoulli distribution with parameter p_{R} and performs recombination if the generated bit evaluates to one.

Let $G = (V, E)$ be a strongly connected directed graph let $w \colon E \to \mathbb{R}^+$ be a weight function on E. The $\mathrm{EA}_{\mathrm{APSP}}$ is the population-based evolutionary algorithm for the APSP problem on (G, w) with recombination probability p_{R}.

Algorithm 5.38 ($\mathrm{EA}_{\mathrm{APSP}}$). *Let the recombination probability p_{R} be given. Let W be the set of all pairs of distinct vertices in V^2. The search space $\mathcal{S}$ of the* $\mathrm{EA}_{\mathrm{APSP}}$ *consists of all directed v-w-paths in G with $(v, w) \in W$. The multi-set of initial search points $x^{(0)}$ contains for each edge $(v, w) \in E$ the path $(v, (v, w), w)$ on this single edge. For two paths P and Q in $\mathcal{S}$, let $P \succeq Q$ if and only if P and Q have the same start-vertex and end-vertex and if $w(P) \le w(Q)$.*

The $\mathrm{EA}_{\mathrm{APSP}}$ *iteratively generates a sequence of multi-sets $\{x^{(t)}\}_{t \in \mathbb{N}}$ of search points in $\mathcal{S}$ by the following procedure.*

For all $t \in \mathbb{N}$ with $t \ge 1$, the population-based evolutionary algorithm generates a random candidate search point $y^{(t)}$. With probability p_{R}, the point $y^{(t)}$ is generated by recombination and with probability $1 - p_{\mathrm{R}}$ by variation. Afterwards, we set

$$x^{(t)} := \begin{cases} \{x \in x^{(t-1)} \colon y^{(t)} \not\succeq x\} \cup \{y^{(t)}\} & \text{if } \forall\, x \in x^{(t-1)} \colon x \not\succeq y^{(t)}, \\ x^{(t-1)} & \text{otherwise.} \end{cases}$$

Recombination

1. *Choose two paths P and Q uniformly at random from $x^{(t-1)}$.*
2. *If $P \circ Q \in \mathcal{S}$ then return $P \circ Q$.*
3. *Otherwise, return either P or Q, each with equal probability.*

Variation

1. *Choose an path $P = (v, \ldots, w)$ uniformly at random from $x^{(t-1)}$.*
2. *Repeat the following steps k times, where k is chosen according to* $\mathrm{Pois}(1)$*:*
   (a) *Choose $u \in \{v, w\}$ uniformly at random.*
   (b) *Choose an edge $e \in E$ incident to u uniformly at random.*
   (c) *If e connects u to P then return P without u and e.*
   (d) *If $u = v$, and $e = (x, v)$ is an incoming edge of v, and $(x, (x, v), v) \circ P \in \mathcal{S}$, then return $(x, (x, v), v) \circ P$.*
   (e) *If $u = w$, and $e = (w, y)$ is an outgoing edge of w, and $P \circ (w, (w, y), y) \in \mathcal{S}$, then return $P \circ (w, (w, y), y)$.*
   (f) *Otherwise, return P*

The *optimization time* T of the $\mathrm{EA}_{\mathrm{APSP}}$ is the first point in time $t \in \mathbb{N}$ for which $x^{(t)}$ contains an optimal path for every $(v, w) \in W$. With probability tending to one, the optimization time of the $\mathrm{EA}_{\mathrm{APSP}}$ is $\mathcal{O}(|V|^{3+1/4} \ln^{1/4} |V|)$.

Theorem 5.39 (Doerr and Theile (2009)). *Let $p_{\mathrm{R}} \in (0, 1)$ be constant. Let the graph $G = (V, E)$ be directed and strongly connected and let $w\colon E \to \mathbb{R}^+$ be a weight function on E. Furthermore, let T be the optimization time of the $\mathrm{EA}_{\mathrm{APSP}}$ with recombination probability $p_{\mathrm{R}} \in (0, 1)$. Then for all $\lambda > 0$ there exist a constant C_λ such that*

$$\Pr\left[T \le C_\lambda\, |V|^{3+1/4} \ln^{1/4} |V|\right] \ge 1 - |V|^{-\lambda}\,.$$

To prove this theorem, we give some preliminary definitions and two propositions. In the following, let $n := |V|$.

Let $(v, w) \in W$. Then $T_{(v,w)}$ is the random variable that describes the first point in time $t \in \mathbb{N}$ for which there is an optimal v-w-path in $x^{(t)}$.

For $g \in \mathbb{R}_0^+$, we call the pair $(x, y) \in W$ a *g-approximation* of (u, w) if there exist an optimal v-w-path containing an x-y-subpath that is at most g edges shorter that the v-w-path. Furthermore, let $T_{g,(v,w)}$ be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $x^{(t)}$ contains an optimal x-y-path such that $(x, y) \in W$ is a g-approximation of (v, w). Thus, $T_{(v,w)} = T_{0,(v,w)}$

Proposition 5.40. *Let $(v, w) \in W$, and $g \in \mathbb{R}_0^+$ with $g \ge 4 \ln n$. Then, for all $\lambda \ge 2$ and n sufficiently large,*

$$\Pr\left[T_{(v,w)} \le T_{g,(v,w)} + \tfrac{6\,\lambda\, g\, n^3}{1-p_{\mathrm{R}}}\right] \ge 1 - n^{1-\lambda}\,.$$

Proof. We may condition on the event that $T_{(v,w)} > T_{g,(v,w)}$ holds since otherwise the event that $T_{(v,w)} \le T_{g,(v,w)} + \frac{6\,\lambda\, g\, n^3}{1-p_{\mathrm{R}}}$ holds with certainty. Thus, without loss of generality, suppose that $T_{(v,w)} > T_{g,(v,w)}$.

For $t \ge T_{g,(v,w)}$, let $h(t)$ be the minimal integer such that $x^{(t)}$ contains an optimal x-y-path for which (x, y) is a $h(t)$-approximation of (v, w). Then, $h(T_{v,w}) = 0$ and also $1 \le h(T_{g,(v,w)}) \le g$.

Next, we define the indicator variables $I_t \in \{0, 1\}$ for all $t \in \mathbb{N}$. For all $t \le T_{(v,w)}$, we let $I_t = 1$ if and only if $h(t) < h(t-1)$. For all $t > T_{(v,w)}$, we let I_t be drawn independently according to the Bernoulli distribution with parameter $\frac{1}{2\,e\,n^3}$, that is, $\Pr[I_t = 1] = \frac{1-p_{\mathrm{R}}}{2\,e\,n^3}$.

Let $T_{g,(v,w)} \le t < T_{(v,w)}$. We show that

$$\Pr\left[I_{t+1} \,\middle|\, I_{T_{g,(v,w)}+1}, \ldots, I_t\right] \ge \frac{1-p_{\mathrm{R}}}{2\,e\,n^3}\,. \tag{5.10.1}$$

By the definition of $h(t)$, there is an optimal v-w-path $(v = u_0, \ldots, u_k = w)$ with $k \in \mathbb{N}$ and two indices $a(t), b(t) \in \{0, \ldots, k\}$ with $a(t) < b(t)$ such that $a(t) + (k - b(t)) = h(t)$ and $x^{(t)}$ contains an optimal $u_{a(t)}$-$u_{b(t)}$-path.

Since $t < T_{(v,w)}$, it holds that $a(t) \ge 1$ or $k - b(t) \ge 1$. Without loss of generality, suppose $a(t) \ge 1$. Then the variation that chooses the optimal $u_{a(t)}$-$u_{b(t)}$-path from $x^{(t)}$ and adds the edge $(u_{a(t)-1}, u_{a(t)})$ decreases $h(t)$.

The probability that this variation happens depends on $|x^{(t)}|$ and the degree of $u_{a(t)}$. However, independently of $x^{(t)}$ (and all $X^{(s)}$ with $s < t$), this probability is at least $\frac{1-p_R}{2\,e\,n^3}$. Thus, (5.10.1) holds.

Now, consider the time interval

$$J := \left\{T_{g,(v,w)} + 1, \ldots, T_{g,(v,w)} + \left\lfloor \tfrac{6\,\lambda\,g\,n^3}{1-p_R} \right\rfloor \right\}.$$

If $\sum_{t\in J} I_t \geq g$, then the event $T_{(v,w)} \leq T_{g,(v,w)} + \frac{6\,\lambda\,g\,n^3}{1-p_R}$ holds with certainty. Thus,

$$\Pr\left[T_{(v,w)} \leq T_{g,(v,w)} + \tfrac{6\,\lambda\,g\,n^3}{1-p_R}\right] \geq \Pr\left[\sum_{t\in J} I_t \geq g\right].$$

By Theorem 4.13, we can now derive the proposition from the Chernoff bound (Theorem 4.2). Let $\mu = \mathrm{E}\big[\sum_{t\in J} I_t\big]$. Then, for sufficiently large n,

$$\mu = \left\lfloor \tfrac{6\,\lambda}{1-p_R}\, g\, n^3 \right\rfloor \cdot \tfrac{1-p_R}{2\,e\,n^3} > \lambda\, g\,.$$

Hence, using the Chernoff bounds we get

$$\Pr\left[\sum_{t\in J} I_t < g\right] < \Pr\left[\sum_{t\in J} I_t < \tfrac{1}{\lambda}\,\mu\right] \leq e^{-\frac{(\lambda-1)^2\,g}{2\lambda}} \leq n^{1-\lambda},$$

which concludes the proof. □

Next, we let W_r be the set of all pairs $(v, w) \in W$ such that there exists an optimal v-w-path on at most r edges. Note that $W_n = W$, since there exists an optimal v-w-path on at most n edges for all $(v, w) \in W$. Furthermore, for all $r \in \mathbb{R}_0^+$ let $T_r := \max_{(v,w)\in W_r} T_{(v,w)}$ be the random variable that describes the first point in time $t \in \mathbb{N}$ for which there is an optimal v-w-path for all $(v, w) \in W_r$ in $x^{(t)}$. Clearly, $T_1 = 0$.

Finally, for $r, g \in \mathbb{N}$ with $g \leq r$ let $T_{g,r} := \max_{(v,w)\in W_r} T_{g,(v,w)}$ be the random variable that describes the first point in time $t \in \mathbb{N}$ such that for all $(v, w) \in W_r$ there is an optimal x-y-path in $x^{(t)}$ where (x, y) is a g-approximation of (v, w). Thus, $T_{r,r} = 0$.

Proposition 5.41. *Let $r, g \in \mathbb{R}_0^+$ with $g \leq r/2$ and $(v, w) \in W_{3r/2}\backslash W_r$. Then, for $\lambda > 0$ and n sufficiently large,*

$$\Pr\left[T_{g,(v,w)} \leq T_r + \tfrac{4\,\lambda\,n^4\,\ln n}{p_R\,r\,g^2}\right] \geq 1 - n^{-\lambda}.$$

Proof. We may condition on the event that $T_{g,(v,w)} > T_r$ holds, otherwise the event $T_{g,(v,w)} \leq T_r + \frac{4\,\lambda\,n^4\,\ln n}{p_R\,r\,g^2}$ holds with certainty. Thus, without loss of generality, suppose that the event $T_{g,(v,w)} > T_r$ holds.

Let $t > T_r$. By the definitions of W_r and $W_{3r/2}$, there exists an optimal v-w-path $P := (v = u_0, (u_0, u_1), \ldots, u_k = w)$ on $k \in \mathbb{N}$ edges such that

$$r < k \leq 3r/2\,.$$

Let $a, b, j \in \mathbb{N}$ with $a + k - b \leq g$, $j \leq a + r$, and $j \geq b - r$. Since P is optimal, the two paths $(v = u_a, \ldots, u_j)$ and $(u_j, \ldots, u_b = w)$ are optimal, too. Moreover, by definition of j both paths are of length at most r. Thus, since $t > T_r$, at time t the $\mathrm{EA}_{\mathrm{APSP}}$ has found an optimal u_a-u_j-path and an optimal u_j-u_b-path. The concatenation of these two optimal paths results in an optimal u_a-u_b-path. Hence, the vertex pair (u_a, u_b) is a g-approximation of (v, w).

Let a, b, and j be given. Then, the probability that $\mathrm{EA}_{\mathrm{APSP}}$ performs a recombination step using this particular concatenation at time t is at least p_{R}/n^4. There are at least $g^2/2$ ways to choose the pair (a, b) and at least $2r - k + 1 \geq r/2$ ways to choose j. Note that these bounds hold independently of the variations and recombinations performed by the $\mathrm{EA}_{\mathrm{APSP}}$ at times $T_r + 1, \ldots, t - 1$. Thus, for all $t > T_r$,

$$\Pr\left[T_{g,(v,w)} = t \,\middle|\, T_{g,(v,w)} \geq t\right] \geq \frac{p_{\mathrm{R}}\, r\, g^2}{4n^4}\,.$$

Let $\Delta := \frac{4\,n^4}{p_{\mathrm{R}}\, r\, g^2}$. Then, by the union bound,

$$\Pr\left[T_{g,(v,w)} \geq T_r + \lambda\,\Delta\,\ln n\right] \leq \left(1 - \tfrac{1}{\Delta}\right)^{\lambda\Delta\ln n} \leq \mathrm{e}^{-\lambda\ln n} = n^{-\lambda}\,,$$

which concludes the proof. □

Using the previous two propositions, we prove Theorem 5.39.

Proof of Theorem 5.39. Let $\lambda > 3$ and let n be sufficiently large. Let $L := \lceil \log_{3/2} n \rceil$ and

$$r(i) := \left(\tfrac{3}{2}\right)^{i} n^{1/4}\,\ln^{1/4} n \qquad \text{and} \qquad g(i) := \left(\tfrac{3}{2}\right)^{-i/3} n^{1/4}\,\ln^{1/4} n$$

for all $i \in \{0, \ldots, L\}$. Then, $g(0) = r(0) = n^{1/4}\,\ln^{1/4} n$ and $g(n) \leq r(n)/2$ for all $n \geq 1$. Furthermore, we have

$$0 = T_{g(0),r(0)} \leq T_{r(0)} \leq T_{g(1),r(1)} \leq T_{r(1)} \leq \cdots \leq T_{g(L),r(L)} \leq T_{k(L)} = T\,.$$

For all $(v, w) \in W_{r(0)}$ let

$$\begin{aligned} A_{(v,w)}\colon &\qquad T_{(v,w)} \leq T_{g(0),(v,w)} + \frac{6\,\lambda\,g(0)\,n^3}{1 - p_{\mathrm{R}}} \quad \text{and} \\ B_{(v,w)}\colon &\qquad T_{g(0),(v,w)} = 0 \qquad \text{(this event occurs with certainty)}\,. \end{aligned}$$

For all $i \in \{1, \ldots, L\}$ and all $(v, w) \in W_{r(i)} \setminus W_{r(i-1)}$ let

$$\begin{aligned} A_{(v,w)}\colon &\qquad T_{(v,w)} \leq T_{g(i),(v,w)} + \frac{6\,\lambda\,g(i)\,n^3}{1 - p_{\mathrm{R}}} \quad \text{and} \\ B_{(v,w)}\colon &\qquad T_{g(i),(v,w)} \leq T_{r(i-1)} + \frac{4\,\lambda\;n^4\,\ln n}{p_{\mathrm{R}}\, r(i-1)\, g(i)^2}\,. \end{aligned}$$

Suppose the events $A_{(v,w)}$ and $B_{(v,w)}$ hold for all $(v, w) \in W$. Then, it holds for all $i \in \{0, \ldots, L\}$, that

$$T_{r(i)} - T_{g(i),r(i)} \leq \frac{6\,\lambda\,g(i)\,n^3}{1 - p_{\mathrm{R}}} = \left(\tfrac{2}{3}\right)^{i/3} \cdot \frac{6\,\lambda\,p_{\mathrm{R}}\,n^{3+1/4}\,\ln^{1/4} n}{p_{\mathrm{R}}\,(1 - p_{\mathrm{R}})}\,,$$

and, for all $i \in \{1, \ldots, L\}$, that

$$T_{g(i),r(i)} - T_{r(i-1)} \leq \frac{4\,\lambda\, n^4\, \ln n}{p_{\mathrm{R}}\, r(i-1)\, g(i)^2} = \left(\tfrac{2}{3}\right)^{i/3} \cdot \frac{6\,\lambda\,(1-p_{\mathrm{R}})\, n^{3+1/4}\, \ln n^{1/4}}{p_{\mathrm{R}}\,(1-p_{\mathrm{R}})}\,.$$

Thus, by the geometric series we have

$$T \leq \tfrac{48}{p_{\mathrm{R}}\,(1-p_{\mathrm{R}})}\,\lambda\, n^{3+1/4}\, \ln n^{1/4}\,.$$

Applying the union bound together with Proposition 5.40 and Proposition 5.41 we derive that $A_{(v,w)}$ and $B_{(v,w)}$ hold for all $(v,w) \in W$ with probability at least $1-2\,n^{3-\lambda}$ which concludes the proof. □

In Theorem 5.39, the recombination probability is restricted to the open interval $(0,1)$. For $p_{\mathrm{R}} = 0$, the $\mathrm{EA}_{\mathrm{APSP}}$ performs variation only. It was shown in Doerr et al. (2008a), that in this case the expected optimization time increases to $\Omega(n^4)$ for the K_n where each edge has weight n except for a Hamilton path with edge-weights 1.

For $p_{\mathrm{R}} = 1$, the $\mathrm{EA}_{\mathrm{APSP}}$ performs recombination only. For this case, it is possible to show an upper bound of $\mathcal{O}(|V|^4 \ln |V|)$ on the expected runtime of the $\mathrm{EA}_{\mathrm{APSP}}$. This is done by setting $p_{\mathrm{R}} = 1$, $g = 0$ and $r = (\frac{3}{2})^i$ in Proposition 5.41 and the summing over all i in $1, \ldots, \lceil \log_{3/2} |V| \rceil$ like in the proof of the previous theorem.

If we take a closer look at Proposition 5.41, it becomes clear that the $|V|^4$ factor in this expected optimization time results from the lower bound on the probability to concatenate two specific paths in the population $x^{(t)}$ at time t. This bound drops to $|V|^3$, if we restrict the selection operator for recombination to pairs of paths such that the end vertex of the first path is the start vertex of the second. In this case the bound on the expected optimization time drops to $\mathcal{O}(|V|^3 \ln |V|)$.

The classical problem-specific algorithm for the APSP is the Floyd-Warshall algorithm (Floyd (1962); Warshall (1962)). This algorithm solves the APSP problem in time $\Theta(|V|^3)$. It uses dynamic programming techniques based on concatenation. Basically, the $\mathrm{EA}_{\mathrm{APSP}}$ using recombination mimics the Floyd–Warshall algorithms. Further studies of this capability of the recombination based evolutionary algorithms to simulate dynamic programming is studied in a general context by Doerr, Eremeev, Horoba, Neumann, and Theile (2009).

5.11. Conclusion

We have analyzed the optimization time of the (1+1) EA for several of the most common polynomial time solvable problems in combinatorial optimization. For some of these problems, we found that the (1+1) EA simulates known problem-specific algorithms. For others, the (1+1) EA does not find an optimal solution in polynomial time with probability tending to one.

The collection of problems we examined is by no means exhaustive. For example, another combinatorial problem that is solvable in polynomial time is the sorting problem. The (1+1) EA solves this problem in polynomial time (Scharnow et al. (2004); Doerr and Happ (2008)).

On the other hand, the min-cut problem is a second example of a polynomial time solvable problem for which the (1+1) EA has exponential expected optimization time (Neumann, Reichel, and Skutella (2008)).

There also exist a number of runtime result for the (1+1) EA on NP-hard problems. He and Yao (2001) studied the the subset-sum problem, Storch (2006, 2007) the maximum clique problem, Neumann (2007) the multi-objective minimum spanning tree problem, and Horoba (2009) the multi-objective shortest path problem.

Again, not all NP-hard problems are equally accessible to the (1+1) EA. For the special case of the partition problem, Witt (2005) showed that the (1+1) EA can find a 4/3-approximation in quadratic time. He also showed that multiple runs of the (1+1) EA will result in a PRAS.

In contrast to this, Friedrich, Hebbinghaus, Neumann, He, and Witt (2007) showed that with probability tending to one, the (1+1) EA for the vertex-cover and set-cover problems does not find a constant factor approximation in polynomial time.

However, under certain conditions this inapproximability can be dealt with by approaches based on multi-objective optimization (Friedrich et al. (2007)), multiple runs (Oliveto, He, and Yao (2007)), the use of diversity mechanisms (Oliveto et al. (2008)), or hybridization (Friedrich, He, Hebbinghaus, Neumann, and Witt (2009)).

For a multi-objective approach, Kratsch and Neumann (2009) conducted a bidimensional analysis using the optimal value as fixed parameter.

Indication of source. The content of this chapter will appear in the book *Theory of Randomized Search Heuristics* (Auger and Doerr (2010)). The chapter contains results that have been previously published in the *Proceedings of GECCO '07* (Doerr and Johannsen (2007b)) and in the *Proceedings of GECCO '10* (Doerr and Johannsen (2010)).

A
Further Contributions

The following contributions to the field of theoretical computer science have not been included in this thesis but are part of my PhD research. Here, I give the abstracts and the references of the respective publications.

Quantum Search Heuristics

Can Quantum Search Accelerate Evolutionary Algorithms?

In this article, we formulate for the first time the notion of a quantum evolutionary algorithm. In fact we define a quantum analogue for any elitist (1+1) randomized search heuristic. The quantum evolutionary algorithm, which we call *(1+1) quantum evolutionary algorithm* (QEA), is the quantum version of the classical (1+1) evolutionary algorithm (EA), and runs only on a quantum computer. It uses Grover search (Grover (1996)) to accelerate the search for improved offsprings.

To understand the speedup of the (1+1) QEA over the (1+1) EA, we study the three well known pseudo-Boolean optimization problems OneMax, LeadingOnes, and Discrepancy. We show that although there is a speedup in the case of OneMax and LeadingOnes in the quantum setting, the speedup is less than quadratic. For Discrepancy, we show that the speedup is at best constant.

The reason for this inconsistency is due to the difference in the probability of making a successful mutation. On the one hand, if the probability of making a successful mutation is large then quantum acceleration does not help much. On the other hand, if the probabilities of making a successful mutation is small then quantum enhancement indeed helps.

In: *Proceedings of GECCO '10* (Johannsen, Kurur, and Lengler (2010)).

SAT for Restricted CNF Formulas

Solving SAT for CNF formulas with a one-sided restriction on variable occurrences
In this paper we consider the class of boolean formulas in Conjunctive Normal Form (CNF) where for each variable all but at most d occurrences are either positive or negative. This class is a generalization of the class of CNF formulas with at most d occurrences (positive and negative) of each variable which was studied in Wahlström (2005).

Applying complement search (Purdom (1984)), we show that for every d there exists a constant $\gamma_d < 2 - \frac{1}{2d+1}$ such that satisfiability of a CNF formula on n variables can be checked in runtime $\mathcal{O}(\gamma_d^n)$ if all but at most d occurrences of each variable are either positive or negative. We thoroughly analyze the proposed branching strategy and determine the asymptotic growth constant γ_d more precisely. Finally, we show that the trivial $\mathcal{O}(2^n)$ barrier of satisfiability checking can be broken even for a more general class of formulas, namely formulas where the positive or negative literals of every variable have what we will call a *d–covering*.

To the best of our knowledge, for the considered classes of formulas there are no previous non-trivial upper bounds on the complexity of satisfiability checking.

In: *Proceedings of SAT '09* (Johannsen, Razgon, and Wahlström (2009)).

Ant Colony Optimization

How Single Ant ACO Systems Optimize Pseudo-Boolean Functions
We undertake a rigorous experimental analysis of the optimization behavior of the two most studied single ant ACO systems on several pseudo-boolean functions. By tracking the behavior of the underlying random processes rather than just regarding the resulting optimization time, we gain additional insight into these systems. A main finding is that in those cases where the single ant ACO system performs well, it basically simulates the much simpler (1+1) Evolutionary Algorithm.

In: *Proceedings of PPSN '08* (Doerr, Johannsen, and Tang (2008b)).

Refined Runtime Analysis of a Basic Ant Colony Optimization Algorithm
Neumann and Witt (2006) analyzed the runtime of the basic ant colony optimization (ACO) algorithm 1-Ant on pseudo-boolean optimization problems. For the problem OneMax they showed how the runtime depends on the evaporation factor. In particular, they proved a phase transition from exponential to polynomial runtime. In this work, we simplify the view on this problem by an appropriate translation of the pheromone model. This results in a profound simplification of the pheromone update rule and, by that, a refinement of the results of Neumann and Witt. In particular, we show how the exponential runtime bound gradually changes to a polynomial bound inside the phase of transition.

In: *Proceedings of CEC '07* (Doerr and Johannsen (2007a)).

Bibliography

Niels H. Abel. Beweis eines Ausdrucks von welchem die Binomial-Formel ein einzelner Fall ist. *Journal für die reine und angewandte Mathematik*, 1:159–160, 1826.

Luca C. Aleardi, Olivier Devillers, and Gilles Schaeffer. Optimal succinct representations of planar maps. In *SCG '06: Proceedings of the 22nd Annual Symposium on Computational Geometry*, pages 309–318. ACM, 2006.

Noga Alon and Joel H. Spencer. *The probabilistic method*. Interscience Series in Discrete Mathematics and Optimization. Wiley, third edition, 2008.

Anne Auger and Benjamin Doerr, editors. *Theory of Randomized Search Heuristics*. World Scientific, 2010. To appear.

Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures and Algorithms*, 19(3-4):194–246, 2001.

Surender Baswana, Somenath Biswas, Benjamin Doerr, Tobias Friedrich, Piyush P. Kurur, and Frank Neumann. Computing single source shortest paths using single-objective fitness. In *FOGA '09: Proceedings of the 10th ACM Workshop on Foundations of Genetic Algorithms*, pages 59–66. ACM, 2009.

Nicla Bernasconi, Konstantinos Panagiotou, and Angelika Steger. On properties of random dissections and triangulations. In *SODA '08: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 132–141. SIAM, 2008.

Hans-Georg Beyer, Hans-Paul Schwefel, and Ingo Wegener. How to analyse evolutionary algorithms. *Theoretical Computer Science*, 287(1):101–130, 2002.

Manuel Bodirsky, Clemens Gröpl, Daniel Johannsen, and Mihyun Kang. A direct decomposition of 3-connected planar graphs. *Séminaire Lotharingien de Combinatoire*, B54Ak:15, 2007.

Peter Cameron, Daniel Johannsen, Thomas Prellberg, and Pascal Schweitzer. Counting defective parking functions. *Electronic Journal of Combinatorics*, 15(1):R92, 2008.

Philippe Chassaing and Guy Louchard. Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures and Algorithms*, 21(1):76–119, 2002.

Robert M. Corless, Gaston H. Gonnet, D. E. G. Hare, David J. Jeffrey, and Donald E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, 1996.

Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Cliff Stein. *Introduction to Algorithms*. MIT Press, second edition, 2001.

Deepak Dhar. Self organized critical state of sandpile automaton models. *Physical Review Letters*, 64(14):1613–1616, 1990.

Reinhard Diestel. *Graph Theory*. Springer, third edition, 2005.

Edsger W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):11–12, 1959.

Benjamin Doerr and Edda Happ. Directed trees: A powerful representation for sorting and ordering problems. In *CEC '08: Proceedings of the 2008 IEEE Congress on Evolutionary Computation*, pages 3606–3613. IEEE, 2008.

Benjamin Doerr and Daniel Johannsen. Refined runtime analysis of a basic ant colony optimization algorithm. In *CEC '07: Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pages 501–507. IEEE, 2007a.

Benjamin Doerr and Daniel Johannsen. Adjacency list matchings — an ideal genotype for cycle covers. In *GECCO '07: Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference*, pages 1203–1210. ACM, 2007b.

Benjamin Doerr and Daniel Johannsen. Edge-based representation beats vertex-based representation in shortest path problems. In *GECCO '10: Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference*, pages 759–766. ACM, 2010.

Benjamin Doerr and Madeleine Theile. Improved analysis methods for crossover-based algorithms. In *GECCO '09: Proceedings of the 11th Annual Genetic and Evolutionary Computation Conference*, pages 247–254. ACM, 2009.

Benjamin Doerr, Edda Happ, and Christian Klein. A tight analysis of the (1+1)-EA for the single source shortest path problem. In *CEC '07: Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pages 1890–1895. IEEE, 2007a.

Benjamin Doerr, Nils Hebbinghaus, and Frank Neumann. Speeding up evolutionary algorithms through asymmetric mutation operators. *Evolutionary Computation*, 15(4):401–410, 2007b.

Benjamin Doerr, Christian Klein, and Tobias Storch. Faster evolutionary algorithms by superior graph representation. In *FOCI '07: Proceedings of the IEEE Symposium on Foundations of Computational Intelligence*, pages 245–250. IEEE, 2007c.

Benjamin Doerr, Edda Happ, and Christian Klein. Crossover can provably be useful in evolutionary computation. In *GECCO '08: Proceedings of the 10th Annual Genetic and Evolutionary Computation Conference*, pages 539–546. ACM, 2008a.

Benjamin Doerr, Daniel Johannsen, and Ching Hoo Tang. How single ant ACO systems optimize pseudo-Boolean functions. In *PPSN '08: Proceedings of the 10th International Conference on Parallel Problem Solving from Nature*, volume 5199 of *Lecture Notes in Computer Science*, pages 378–388. Springer, 2008b.

Benjamin Doerr, Anton Eremeev, Christian Horoba, Frank Neumann, and Madeleine Theile. Evolutionary algorithms and dynamic programming. In *GECCO '09: Proceedings of the 11th Annual Genetic and Evolutionary Computation Conference*, pages 771–778. ACM, 2009.

Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Drift analysis and linear functions revisited. In *CEC '10: Proceedings of the 2010 IEEE Congress on Evolutionary Computation*. IEEE, 2010a. To appear.

Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative drift analysis. In *GECCO '10: Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference*, pages 1449–1456. ACM, 2010b.

Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276(1–2):51–81, 2002.

Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4–5):577–625, 2004.

Ioana Dumitriu, Joel H. Spencer, and Catherine H. Yan. Branching processes with negative offspring distributions. *Annals of Combinatorics*, 7(1):35–47, 2003.

Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8(1):30–32, 1741.

Pasquale J. Federico. The number of polyhedra. *Philips Research Reports*, 30(1):220–231, 1975.

William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, third edition, 1968.

Simon Fischer and Ingo Wegener. The Ising model on the ring: mutation versus recombination. In *GECCO '04: Proceedings of the 6th Annual Genetic and Evolutionary Computation Conference, Part I*, volume 3102 of *Lecture Notes in Computer Science*, pages 1113–1124. Springer, 2004.

Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, first edition, 2009.

Robert W. Floyd. Algorithm 97: shortest path. *Communications of the ACM*, 5(6):1–13, 1962.

Dominique Foata and John Riordan. Mappings of acyclic and parking functions. *Aequationes Mathematicae*, 10(1):10–22, 1974.

Philippe Di Francesco. Private communication, 2008.

Tobias Friedrich, Nils Hebbinghaus, Frank Neumann, Jun He, and Carsten Witt. Approximating covering problems by randomized search heuristics using multi-objective models. In *GECCO '07: Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference*, pages 797–804. ACM, 2007.

Tobias Friedrich, Jun He, Nils Hebbinghaus, Frank Neumann, and Carsten Witt. Analyses of simple hybrid algorithms for the vertex cover problem. *Evolutionary Computation*, 17(1): 1006–1029, 2009.

Éric Fusy. Quadratic exact-size and linear approximate-size random generation of planar graphs. In *AofA '05: Proceedings of the 2005 International Conference on Analysis of Algorithms*, pages 125–138. DMTCS, 2005.

Éric Fusy, Dominique Poulalhon, and Gilles Schaeffer. Dissections and trees, with applications to optimal mesh encoding and to random sampling. In *SODA '05: Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 690–699. SIAM, 2005.

Zhicheng Gao and Nicholas C. Wormald. Sharp concentration of the number of submaps in random planar triangulations. *Combinatorica*, 23(3):467–486, 2003.

Zhicheng Gao and Nicholas C. Wormald. The size of the largest components in random planar maps. *SIAM Journal on Discrete Mathematics*, 12(2):217–228, 1999.

Oliver Giel and Per Kristian Lehre. On the effect of populations in evolutionary multi-objective optimization. In *GECCO '06: Proceedings of the 8th Annual Genetic and Evolutionary Computation Conference*, pages 651–658. ACM, 2006.

Oliver Giel and Ingo Wegener. Evolutionary algorithms and the maximum matching problem. In *STACS '03: Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science*, volume 2607 of *Lecture Notes in Computer Science*, pages 415–426. Springer, 2003.

Oliver Giel and Ingo Wegener. Maximum cardinality matchings on trees by randomized local search. In *GECCO '06: Proceedings of the 8th Annual Genetic and Evolutionary Computation Conference*, pages 539–546. ACM, 2006.

Julian D. Gilbey and Louis H. Kalikow. Parking functions, valet functions and priority queues. *Discrete Mathematics*, 197–198(1):351–373, 1999.

Lov K. Grover. A fast quantum mechanical algorithm for database search. In *STOC '96: Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 212–219. ACM, 1996.

Mark D. Haiman. Conjectures on the quotient ring by diagonal invariants. *Journal of Algebraic Combinatorics*, 3(1):17–76, 1994.

Bruce Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):387–403, 1982.

Edda Happ, Daniel Johannsen, Christian Klein, and Frank Neumann. Rigorous analyses of fitness-proportional selection for optimizing linear functions. In *GECCO '08: Proceedings of the 10th Annual Genetic and Evolutionary Computation Conference*, pages 953–960. ACM, 2008.

Jun He and Xin Yao. Drift analysis and average time complexity of evolutionary algorithms. *Acta Informatica*, 127(1):51–81, 2001.

Jun He and Xin Yao. From an individual to a population: an analysis of the first hitting time of population-based evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 6(5):495–511, 2002.

Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.

Carl Hierholzer. Über die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechung zu umfahren. *Mathematische Annalen*, 6(1):57–63, 1873.

Christian Horoba. Analysis of a simple evolutionary algorithm for the multiobjective shortest path problem. In *FOGA '09: Proceedings of the 10th ACM Workshop on Foundations of Genetic Algorithms*, pages 113–120. ACM, 2009.

Christian Horoba and Dirk Sudholt. Running time analysis of ACO systems for shortest path problems. In *SLS '09: Proceedings of the 2nd International Workshop on Engineering Stochastic Local Search Algorithms*, volume 5752 of *Lecture Notes in Computer Science*, pages 76–91. Springer, 2009.

Thomas Jansen and Dirk Sudholt. Analysis of an asymmetric mutation operator. *Evolutionary Computation*, 18(1):1–26, 2010.

Jarník. O jistém problému minimálním. *Práca Moravské Přírodovědecké Společnosti*, 6(1): 48–50, 1930.

Daniel Johannsen and Konstantinos Panagiotou. Vertices of degree k in random maps. In *SODA '10: Proceedings of the 21th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1436–1447. SIAM, 2010.

Daniel Johannsen, Igor Razgon, and Magnus Wahlström. Solving SAT for CNF formulas with a one-sided variable occurrence restriction. In *SAT '09: Proceedings of the 12th International Conference on Theory and Applications of Satisfiability Testing*, volume 5584 of *Lecture Notes in Computer Science*, pages 80–85, 2009.

Daniel Johannsen, Piyush P. Kurur, and Johannes Lengler. Can quantum search accelerate evolutionary algorithms? In *GECCO '10: Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference*, pages 1433–1440. ACM, 2010.

Alan G. Konheim and Benjamin Weiss. An occupancy discipline and applications. *SIAM Journal on Applied Mathematics*, 14(6):1266–1274, 1966.

Stefan Kratsch and Frank Neumann. Fixed-parameter evolutionary algorithms and the vertex cover problem. In *GECCO '09: Proceedings of the 11th Annual Genetic and Evolutionary Computation Conference*, pages 293–300. ACM, 2009.

Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):1389–1401, 1956.

Joseph P. S. Kung and Catherine H. Yan. Exact formula for moments of sums of classical parking functions. *Advances in Applied Mathematics*, 31(1):215–241, 2003a.

Joseph P. S. Kung and Catherine H. Yan. Končarov polynomials and parking functions. *Journal of Combinatorial Theory, Series A*, 102(1):16–37, 2003b.

Valery A. Liskovets. A pattern of asymptotic vertex valency distributions in planar maps. *Journal of Combinatorial Theory,* Series B, 75(1):116–133, 1999.

Satya N Majumdar and David S. Dean. Exact solution of a drop-push model for percolation. *Physical Review Letters*, 89(11):115701.1–115701.4, 2002.

Colin McDiarmid. On method of bounded differences. *Surveys in Combinatorics*, 141(1): 148—188, 1989.

Kurt Mehlhorn and Peter Sanders. *Algorithms and Data Structures: The Basic Toolbox*. Springer, first edition, 2009.

Silvio Micali and Vijay V. Vazirani. An $O(\sqrt{|V|}\,|E|)$ algorithm for finding maximum matching in general graphs. In *FOCS '80: Proceedings of the 21st Annual IEEE Syposium on Foundations of Computer Science*, pages 17–27. IEEE, 1980.

Frank Neumann. Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spanning tree problem. *European Journal of Operational Research*, 181(3):114–131, 2007.

Frank Neumann. Expected runtimes of evolutionary algorithms for the Eulerian cycle problem. *Computers and Operation Research*, 35(9):3–19, 2008.

Frank Neumann and Ingo Wegener. Minimum spanning trees made easier via multi-objective optimization. In *GECCO '05: Proceedings of the 7th Annual Genetic and Evolutionary Computation Conference*, pages 763–769. ACM, 2005.

Frank Neumann and Ingo Wegener. Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. *Theoretical Computer Science*, 378(1):32–40, 2007.

Frank Neumann and Carsten Witt. Runtime analysis of a simple ant colony optimization algorithm. In *ISAAC '06: Proceedings of the 17th International Symposium on Algorithms and Computation*, volume 4288 of *Lecture Notes in Computer Science*, pages 618–627. Springer, 2006.

Frank Neumann, Joachim Reichel, and Martin Skutella. Computing minimum cuts by randomized search heuristics. In *GECCO '08: Proceedings of the 10th Annual Genetic and Evolutionary Computation Conference*, pages 779–786. ACM, 2008.

Pietro S. Oliveto and Carsten Witt. Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica*, 2010. In press.

Pietro S. Oliveto, Jun He, and Xin Yao. Analysis of population-based evolutionary algorithms for the vertex cover problem. In *CEC '08: Proceedings of the 2008 IEEE Congress on Evolutionary Computation*, pages 1563–1570. IEEE, 2008.

Pietro Simone Oliveto, Jun He, and Xin Yao. Evolutionary algorithms and the vertex cover problem. In *CEC '07: Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pages 1870–1877. IEEE, 2007.

Alois Panholzer. Private communication, 2008.

Christos H. Papadimitriou, Alejandro A. Schäffer, and Mihalis Yannakakis. On the complexity of local search. In *STOC '90: Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, pages 438–445. ACM, 1990.

Jim Pitman and Richard P. Stanley. A polytope related to empirical distributions, plane trees, parking functions and the associahedron. *Discrete and Computational Geometry*, 27(4): 603–634, 2002.

Alexander Postnikov and Boris Shapiro. Trees, parking functions, syzygies, and deformations of monomial ideals. *Transactions of the American Mathematical Society*, 356(8):3109–3142, 2004.

Robert C. Prim. Shortest connection networks and some generalizations. *Bell System Technology Journal*, 36(1):269–271, 1957.

Helmut Prodinger. The kernel method: a collection of examples. *Séminaire Lotharingien de Combinatoire*, B50f:19, 2004.

Paul W. Purdom. Solving satisfiability with less searching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(4):510–513, 1984.

Joachim Reichel and Martin Skutella. Evolutionary algorithms and matroid optimization problems. In *GECCO '07: Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference*, pages 947–954. ACM, 2007.

Joachim Reichel and Martin Skutella. On the size of weights in randomized search heuristics. In *FOGA '09: Proceedings of the 10th ACM Workshop on Foundations of Genetic Algorithms*, pages 21–28. ACM, 2009.

L. Bruce Richmond and Nicholas C. Wormald. Almost all maps are asymmetric. *Journal of Combinatorial Theory*, Series B, 63(1):1–7, 1995.

Galen H. Sasaki and Bruce Hajek. The time complexity of maximum matching by simulated annealing. *Journal of the ACM*, 35(2):57–85, 1988.

Jens Scharnow, Karsten Tinnefeld, and Ingo Wegener. The analysis of evolutionary algorithms on sorting and shortest paths problems. *Journal of Modelling and Algorithms*, 3(4):281–293, 2004.

Joel H. Spencer and Catherine H. Yan. An enumeration problem and branching processes. Preprint, 2001.

Richard P. Stanley. Parking functions and noncrossing partitions. *Electronic Journal of Combinatorics*, 4(2):R20, 1997.

Ernst Steinitz. Polyeder und Raumeinteilungen. *Enzyklopedie der mathematischen Wissenschaften*, 3(9):1–139, 1922.

Tobias Storch. How randomized search heuristics find maximum cliques in planar graphs. In *GECCO '06: Proceedings of the 8th Annual Genetic and Evolutionary Computation Conference*, pages 567–574. ACM, 2006.

Tobias Storch. Finding large cliques in sparse semi-random graphs by simple randomized search heuristics. *Theoretical Computer Science*, 386(1–2):32–40, 2007.

William T. Tutte. A census of planar maps. *Canadian Journal of Mathematics*, 15(1):249–271, 1963.

Magnus Wahlström. Faster exact solving of SAT formulae with a low number of occurrences per variable. In *SAT '05: Proceedings of the 8th International Conference on Theory and Applications of Satisfiability Testing*, volume 3569, pages 309–323. Lecture Notes in Computer Science, 2005.

Stephen Warshall. A theorem on Boolean matrices. *Journal of the ACM*, 9(1):11–12, 1962.

Richard A. Watson and Thomas Jansen. A building-block royal road where crossover is provably essential. In *GECCO '07: Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference*, pages 1452–1459. ACM, 2007.

Hassler Whitney. Congruent graphs and the connectivity of graphs. *American Journal of Mathematics*, 54(1):150–168, 1932.

Carsten Witt. Worst-case and average-case approximations by simple randomized search heuristics. In *STACS '05: Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science*, volume 3404 of *Lecture Notes in Computer Science*, pages 44–56. Springer, 2005.

Catherine H. Yan. Generalized parking functions, tree inversions, and multicolored graphs. *Advances in Applied Mathematics*, 27(2–3):641–670, 2001.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, 7 April 2010

Daniel Johannsen