Limit Laws for Random Recursive Structures and Algorithms

Dissertation
zur Erlangung des Doktorgrades
der Mathematischen Fakultät
der Albert-Ludwigs-Universität
Freiburg im Breisgau

vorgelegt von
Ralph Neininger

Oktober 1999

Dekan: Prof. Dr. W. Soergel

1. Referent: Prof. Dr. L. Rüschendorf

2. Referent: Prof. Dr. L. Devroye, Montreal

Datum der Promotion: 10. Dezember 1999

Contents

Introduction

Mathematical Analysis of Algorithms is the field where characteristic parameters of algorithms are studied under a suitable probabilistic model. Most significant parameters of an algorithm are its running time and the amount of storage needed. The stochastic component arises by modeling the (unknown) input by some probability distribution. This distribution is usually chosen uniformly on the set of possible inputs. Also the algorithm itself may be random. One motivation to consider such random algorithms is that their performance then often is independent of a special fixed input. The parameters indicating the performance of the algorithm in a probabilistic model become random variables.

The most fundamental algorithms deal with problems arising in sorting, searching, selection, arithmetic operations, random number generation, and the organization of storage. An encyclopedic treatise are the three volumes of D.E. Knuth (1997a, 1997b, 1998). Such basic algorithms are formulated independently of a specific programming language, so that an analysis does not depend on a particular implementation. One of the most famous algorithms is the Quicksort algorithm created by C.A.R. Hoare in 1961 for sorting a file of items. Quicksort is of great practical interest. A median-of-three variant has become the basis for the Unix "sort" feature. Quicksort for decades has served as a model for the analysis of algorithms in general, since it embodies two key paradigms of the design of algorithms, namely the concept of divide and conquer and randomization.

The mean running time of Quicksort to sort a file of n items, which are permuted uniformly at random, is of the order $\Theta(n \log n)$. However, it is also known that in the worst case the algorithm needs $\mathcal{O}(n^2)$ steps. There exist sorting algorithms which also in the worst case only need $\Theta(n \log n)$ steps. Nevertheless these algorithms are in practice usually beaten by Quicksort. Thus one needs a finer stochastic analysis to gain a more detailed understanding of this behavior. Therefore, in addition to the average case behavior of an algorithm also distributional properties are of interest. This leads to the analysis of higher order moments, exponential moments, concentration around the mean, large deviations, limit laws, and the study of the tails of the distributions. Investigations of this type are the subject of the present work.

The first chapter is devoted to searching. An analysis of the cost of a partial match query in comparison based structures is given. Partial match query is a fundamental search routine in the use of data bases. The structures under consideration are the classical K-d tree, the locally balanced K-d-t tree, the random relaxed K-d tree, and the quadtree. For all these structures first order asymptotics are known for the mean of the cost of a partial match query in the uniform probabilistic model. The variances were derived for the two-dimensional quadtree and the random relaxed K-d tree. In the first chapter the missing variances are derived and limit laws of the scaled costs are given. Furthermore, results on the existence and convergence of the moments and on concentration around the mean are given.

Generalizations of the sorting algorithm Quicksort are discussed in the second chapter. By a well-known equivalence the running time of Quicksort is distributed as the internal path length of the random binary search tree. Parameters of trees like the depth of a node, the height and the internal path length correspond to the costs of insertion operations in a tree. These parameters have been analyzed for various special trees. L. Devroye (1998) introduced a general tree model which includes many common trees and studied the depth of insertion and the height of his random split tree resulting in a uniform validity. In the second chapter an analysis for the internal path length of the random split tree is presented. Under proper assumptions the first order asymptotic of the variance, the limit theorem, and results on exponential moments and large deviations of the internal path length are given. This particularly applies to the random quadtree and the m-ary search tree, for which the limit theorems have been unknown so far.

The third chapter is concerned with the subject of selection. A limit law for the running time of the median-of-three version of the algorithm multiple Quickselect is given. Multiple Quickselect is a generalization of C.A.R. Hoare's Find algorithm. More explicit results for the median-of-three version of Find are stated including the asymptotics of all moments, Laplace and Fourier transforms and large deviations.

The investigations of the first three chapters are based on the contraction method. This method was introduced by U. Rösler (1991) for the derivation of the limit theorem for the running time of Quicksort. The contraction method was further developed independently in Rösler (1992) and Rachev and Rüschendorf (1995). A survey was given in Rösler and Rüschendorf (1999). Applying this method one starts with a distributional recursive equation satisfied by the cost (respectively running time) under consideration. The scaling of the cost leads to a modified recursion for the normalized cost. This modified recursion should converge to a limiting form in a certain way. Then this limiting equation gives rise to a corresponding operator on the space of probability measures being endowed with a metric which is complete on an appropriate subspace. Showing contractivity of the operator and using Banach's fixed point theorem we are led to a fixed point, which is the candidate for the weak limit of the scaled cost. Deriving convergence is the last and technically most intricate step of the method.

The last two chapters are devoted to the analysis of related random recursive structures. In the fourth chapter the point to which an interval splitting scheme shrinks is under consideration. A new type of convergence rate for such splitting schemes is introduced. A relation to products of two-dimensional random stochastic matrices also leads to an approach to the convergence of these products in terms of probability metrics.

In the last chapter a random affine recursion of a branching type is discussed. Limit laws and formulae for the derivation of the first and second moment necessary for the normalization are given. One approach is based on contraction arguments involving L_2-assumptions. In the last section another approach is discussed, which makes use of representations involving products of independent matrices in connection with the concept of Lyapunov exponents.

Acknowledgments

I would like to thank my advisor Professor Ludger Rüschendorf for drawing my attention to the subject of the analysis of algorithms. He continuously gave me support while allowing me also great freedom for own investigations. Further experts of the field also had great influence on this work. Professor Philippe Flajolet exposed to me the problem of partial match query and pointed out its connection to the contraction method which led to the first chapter. Also in the initial state of the second chapter for the analysis of the quadtree Professor Flajolet gave an essential hint to the limiting equation. Professor Luc Devroye suggested that an extension of the quadtree case to the random split tree should be possible. This led to the final version of the second chapter. Last but not least I would like to thank Professor Uwe Rösler for fundamental discussions concerning the contraction method.

Notation

The following notation is used throughout this work. $M^1(\mathbb{R}^d, \mathcal{B}^d)$ stands for the space of probability measures on $\mathbb{R}^d$. By $\mathbb{E}\,\mu$ the expectation of a random variable (r.v.) with distribution $\mu \in M^1(\mathbb{R}^d, \mathcal{B}^d)$ is denoted. For $d = 1$ we use also $\mathrm{Var}\ \mu$ for the corresponding variance. Define

$$M^d_{\gamma,p} := \{\, \mu \in M^1(\mathbb{R}^d, \mathcal{B}^d) \;:\; \mathbb{E}\,\mu = \gamma,\ \int \|x\|^p \, d\mu(x) < \infty \} \tag{0.1}$$

for $\gamma \in \mathbb{R}^d$ and $p \geq 1$. In particular $M^d_{0,2}$ are the centered probability measures on $\mathbb{R}^d$ with existing second moment. In dimension $d = 1$ we abriviate $M_{\gamma,p} := M^1_{\gamma,p}$.

Convergence in probability is denoted by $\stackrel{\mathbb{P}}{\to}$, furthermore $\stackrel{\mathcal{D}}{=}$ and $\sim$ mean equality in distribution either for two random variables or a random variable and a probability measure. The distribution of an r.v. X is denoted by $\mathbb{P}^X$ and $\mathcal{L}(X)$. We write $\mathbb{P}^{X\,|\,Y}$ for the conditional distribution for X given Y.

$B(n,p)$ and $M(n,u)$ are the binomial respectively multinomial distributions with parameters $n \in \mathbb{N}, p \in [0,1]$ and $u \in \mathbb{R}^d$ with $\sum u_i = 1$. If $U = (U_1, \dots, U_d)$ is a random vector with $\sum U_i = 1$ then $X \sim M(n,U)$ states that $\mathbb{P}^{X\,|\,U=u} = M(n,u)$ for $\mathbb{P}^U$–almost all u. The density of the beta distribution beta(a,b) with parameters $a, b > 0$ is

$$f(x) = \mathbf{1}_{[0,1]}(x)\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1} \quad \text{for} \quad x \in \mathbb{R}. \tag{0.2}$$

For $\mu, \nu \in M^1(\mathbb{R}^d, \mathcal{B}^d)$ the convolution is denoted by $\mu * \nu$, the product measure by $\mu \otimes \nu$. If T is an appropriate measurable map then $T(\mu)$ stands for the image measure of μ under T. By $\frac{d\mu}{d\nu}$ the Radon-Nikodym derivative of μ with respect to ν is denoted, λ^d is the d-dimensional Lebesgue measure.

The minimal ℓ_p-metric

$$\ell_p(\mu,\nu) := \inf\{(\mathbb{E}\,\|X-Y\|^p)^{1/p} \,:\, X \stackrel{\mathcal{D}}{=} \mu, Y \stackrel{\mathcal{D}}{=} \nu\} \tag{0.3}$$

is defined for all $p \geq 1$ and $\mu, \nu \in M^1(\mathbb{R}^d, \mathcal{B}^d)$ with existing pth moment. $(M^d_{0,p}, \ell_p)$ is a complete metric space and convergence in the ℓ_p-metric is equivalent to weak convergence plus convergence of the pth moments (cf. Rachev (1991)). For r.v. X, Y also the notation $\ell_p(X,Y) := \ell_p(\mathbb{P}^X, \mathbb{P}^Y)$ is used.

By $\langle\,\cdot\,,\,\cdot\,\rangle$ the standard inner product on $\mathbb{R}^d$ is denoted, $\|\cdot\|$ stands for the euclidian norm either on $\mathbb{R}^d$ or on some set of real matrices, $\|\cdot\|_{op}$ denotes the operator norm of a matrix.

Chapter 1

Partial match query

Databases for multidimensional data are of special interest for many applications in computer science, e.g. for geographical information systems, computer graphics and computational geometry. Data structures for multiattribute keys should support the usual dictionary operations as well as some *associative queries*. Examples for such associative queries are nearest neighbor queries, partial match queries and convex or orthogonal range queries. For each of these queries a certain condition is given and all keys of the file have to be retrieved, which satisfy this condition, for example report all data in the file lying in a certain range of the data space. Data structures which maintain multidimensional data are considered in the books of Knuth (1998) and Samet (1990). These structures can be divided into *comparison based algorithms* and methods based on *digital techniques*. The digital techniques use binary representations of the keys. Examples are *tries*, *digital search trees*, and *Patricia tries*. Examples for comparison based structures are *quadtrees* and *multidimensional binary search trees* (*K-d trees*). These algorithms work with comparisons of the whole keys instead of binary representations. For a stochastic analysis of the performance of basic parameters for these structures see Mahmoud (1992).

In this chapter an asymptotic distributional analysis for the cost of partial match queries in comparison based structures is presented.

We assume the data to belong to some d-dimensional domain $D = D_1 \times \ldots \times D_d$, which using binary encodings we can assimilate to the unit cube $[0,1]^d$. For a partial match query a *query* $q = (q_1, \ldots, q_d)$ is given where $q_i \in [0,1] \cup \{*\}$ for $1 \leq i \leq d$. Here $*$ denotes that this component is left unspecified. Then all data in the file have to be retrieved, which match the query q. This means to report all keys which are identical to q in all the components where q is specified, i.e. the components with $q_i \neq *$. The *specification pattern* $u \in \{S, *\}^K$ of a query q is the vector whose entries are S in the components where the query is specified and $*$ otherwise.

For the probabilistic analysis of partial match retrieval we assume the *uniform* *probabilistic model* following Flajolet and Puech (1986). The uniform probabilistic model assumes all components in the data and the specified components in the query to be independent and uniformly distributed on $[0, 1]$. For comparison based algorithms this is equivalent to the more general model where the components are assumed to be drawn independently from any continuous distribution over any interval. Quadtrees and K-d trees built up by independent and uniformly distributed data are called *random* quadtrees respectively *random* K-d trees.

The quadtree structure is due to Finkel and Bentley (1974). It extends the classical idea of binary search trees to multidimensional data. For the construction of the quadtree we refer to Mahmoud (1992). Essentially a data point partitions the search space by the hyperplanes perpendicular to the axes. Used recursively this principle leads to a decomposition of the search space into quadrants. The quadtree corresponds to this partitioning.

The K-dimensional binary search tree, or K-d tree, was introduced by Bentley in 1975. It is a binary tree in which each node contains a K-dimensional key. Here a data point partitions the search space by a hyperplane perpendicular to one of the axes into two halfspaces. The precise way this is done depends on the special kind of K-d tree under consideration. Note that for quadtrees the dimension of the search space is denoted by d whereas for the K-d trees the dimension is denoted by K.

In this chapter the cost of a partial match query in multidimensional quadtrees, K-d trees and two variants of K-d trees, the locally balanced K-d-t tree of Cunto, Lau and Flajolet (1989) and the random relaxed K-d tree of Duch, Estivill-Castro and Martínez (1998) is investigated. Many further variants of Bentley's original K-d tree have been introduced and analyzed e.g. optimized K-d trees, dynamically balanced K-d trees, divided K-d trees and squarish K-d trees. For references to these trees see the preprints of Duch et al. (1998), Martínez et al. (1998) and Devroye et al. (1999). A limit theorem for the cost of a partial match query in the 2-dimensional trie is given in the preprint of Schachinger (1999).

For a partial match query in all the trees under consideration we have to start at the root of the tree. According to the comparisons of the specified components of the query with the corresponding components of the root some of the subtrees of the root have to be considered recursively for the further search. The cost of a partial match query is measured by the number of nodes traversed during the search. We denote this cost in a tree containing n nodes by C_n.

For the trees treated here the mean of the cost (in the uniform probabilistic model) is known to be

$$\mathbb{E}\, C_n \sim \gamma n^{\alpha-1} \tag{1.1}$$

with some constant $\gamma > 0$ and α in the exponent given by some algebraic equation. These parameters depend on the specification pattern of the query and of course on the special tree. The average performance of quadtrees and K-d trees does not attain the *optimal* order of magnitude of a fully balanced binary tree

$$\sim \text{ const } n^{1-s/K} \tag{1.2}$$

for a query with $1 \leq s \leq K-1$ components of a K-dimensional space specified. This was observed first by Flajolet and Puech (1986) for the case of the K-d tree disproving an older conjecture that random K-d trees behave in the average as fully balanced binary trees. Also K-d-t trees and random relaxed K-d trees do not attain the optimal exponent. Recently Devroye, Jabbour, and Zamora-Cura (1999) obtained the optimal time bound introducing the *squarish* K-d tree, which reaches the average time performance for partial match query of $\Theta(n^{1-s/K})$.

The standard deviation in the case of quadtrees and K-d trees is of the same order of magnitude as the mean. The main results in this chapter are limit laws for the normalized cost

$$X_n := \frac{C_n - \mathbb{E}\, C_n}{n^{\alpha-1}}. \tag{1.3}$$

In each tree X_n converges weakly to a random variables which is characterized as the fixed point of a random affine operator. Explicit first order asymptotics of the variance of C_n are also derived, known so far only for the random relaxed K-d tree and the 2-dimensional quadtree (see Martínez et al. (1998)). At the end of this chapter results on the Laplace transform of X_n are discussed.

For the proofs we use the *contraction method*. From the point of view of the contraction method the problem of partial match query has some similarity to the running time of the Find-algorithm in the model of Mahmoud, Modarres, and Smythe (1995); see chapter 3 of the present work. The fact that mean and standard deviation are of the same order of magnitude simplifies the analysis for these problems compared to the analysis of Quicksort and the related problems of internal path lengths in random trees treated in chapter 2. For the partial match query problem a first order asymptotic of the mean is sufficient in order to define the corresponding *limiting operators*. Nevertheless these limiting operators are more involved than the corresponding operators for the Find-algorithm, which is caused by the purely one-sided character of the Find problem.

1.1 Bentley's K-d tree

The K-d tree is generated by inserting the first key into the root of the initially empty tree. Then the first component of the root serves as a discriminator for the further insertions process. The first components of the following keys are compared with the first component of the root. If they are smaller then they are recursively inserted into the left subtree of the root, otherwise they are inserted recursively into the right subtree. Then on the second level of the tree the second components of the keys are used for the splitting process. On the following levels the components $1, \ldots, K$ are drawn cyclically as discriminators. A detailed description is given in Mahmoud (1992).

A partial match query in a K-d tree starts at the root. At each node the search has to inspect one or both of the subtrees according whether the corresponding component is specified or not. Let $u \in \{S, *\}^K$ be a specification pattern and denote by U, Y the first component of the root respectively the first component of the query if this component is specified. Then in the uniform probabilistic model U, Y are independent and uniformly distributed on $[0, 1]$. The subtrees of a K-d tree are given their cardinality again shaped like binary search trees and mutually independent. This implies that after inspecting the root the search algorithm calls recursively partial match queries in the subtrees in the same probabilistic model. Of course the specification pattern for the subsequent queries has to be shifted left cyclically. Denote the cost of a partial match query with specification pattern u in a K-d tree with n keys inserted by $C_n^{(u)}$. Then with $v \in \{S, *\}^{K-1}$ in distribution the following recursive equations are valid

$$C_n^{(Sv)} \stackrel{\mathcal{D}}{=} \mathbf{1}_{\{Y<U\}} C_Z^{(vS)} + \mathbf{1}_{\{Y \geq U\}} \overline{C}_{n-1-Z}^{(vS)} + 1, \tag{1.4}$$

$$C_n^{(*v)} \stackrel{\mathcal{D}}{=} C_Z^{(v*)} + \overline{C}_{n-1-Z}^{(v*)} + 1. \tag{1.5}$$

Here $(\overline{C}_i^{(vS)}) \sim (C_i^{(vS)})$ and $(\overline{C}_i^{(v*)}) \sim (C_i^{(v*)})$. Z is the cardinality of the left subtree of the root, i.e.

$$\mathbb{P}^{Z \mid U=p} = B(n-1, p) \quad \text{for} \quad p \in [0, 1], \tag{1.6}$$

and $Y, (U, Z), (C_i^{(vS)}), \overline{C}_i^{(vS)})$ respectively $Y, (U, Z), (C_i^{(v*)}), (\overline{C}_i^{(v*)})$ are independent.

We want to show weak convergence for scaled versions of $C_n^{(u)}$. In order to apply the contraction method the equations (1.4) and (1.5) have to be scaled. The distributions (even the expectations) of the $C_n^{(u)}$ depend on the particular specification pattern u. For this reason it is not suitable to scale directly the equations (1.4), (1.5). Let the query inspect the levels $0, \ldots, K-1$ explicitly. Then 2^{K-s} of the 2^K subtrees on level K have to be inspected recursively with the original specification pattern. $1 \leq s \leq K-1$ denotes the number of specified components in the specification pattern u and $1 \leq r_1 < r_2 < \ldots < r_s \leq K$ the coordinates of the specified components. Assume the specification pattern to be fixed. The dependence on u is suppressed in the notation. Denote

$$D_n := \{0, 1\}^n \quad \text{for} \quad n \in \mathbb{N},$$

$$D_0 := \{\emptyset\}, \quad \text{and} \quad D^{(K)} := \bigcup_{n=0}^{K-1} D_n. \tag{1.7}$$

For $\sigma \in D_n, \sigma = (\sigma_1, \ldots, \sigma_n)$ and $1 \leq j \leq n$ let $\sigma|j := (\sigma_1, \ldots, \sigma_j) \in D_j$ and $\sigma|0 := \emptyset$, $|\sigma|$ denotes the length of a $\sigma \in D^{(K)}$. The nodes of the levels $0, \ldots, K-1$ are numbered by the elements of $D^{(K)}$. The nodes on level $0 \leq j \leq K-1$ are counted by the elements of D_j from the left to right in increasing order interpreting the elements of D_j as dual representations of integers. The subtrees on the level K are numbered by the elements of D_K analogously. Assume for a moment that all nodes of the levels $0, \ldots, K-1$ are internal nodes of the K-d tree. Denote by $u_\sigma \in \mathbb{R}$ the component of the key stored in node $\sigma \in D^{(K)}$, which is used as the discriminator for the splitting process (this is the $|\sigma|+1$st component of the key) and by $y_1, \ldots, y_s \in \mathbb{R}$ the specified components of the query. Then for the recursion step exactly those subtrees $\sigma \in D_K$ on level K have to be inspected for the subsequent search, which satisfy

$$\begin{aligned} y_j &< u_{\sigma|(r_j-1)} \quad \text{if} \quad \sigma_{r_j} = 0 \\ y_j &\geq u_{\sigma|(r_j-1)} \quad \text{if} \quad \sigma_{r_j} = 1. \end{aligned} \tag{1.8}$$

for all $1 \leq j \leq s$.

For the stochastic analysis in the uniform probabilistic model denote by $U_\sigma, \sigma \in D^{(K)}$ analogously to u_σ the (random) component of the key in node $\sigma \in D_K$, which is used as discriminator. It has to be distinguished whether the levels $0 \ldots, K-1$ are totally full with keys or not. The probability

$$p_n \in [0,1] \tag{1.9}$$

for the event that the levels $0, \ldots, K-1$ are full after inserting n keys into the empty tree satisfies $p_n \to 1$ for $n \to \infty$. For deeper results concerning the *saturation level* see Devroye (1986). In the case of full levels denote by

$$\mathcal{U}_K := \{U_\sigma : \sigma \in D^{(K)}\} \tag{1.10}$$

this family of discriminators. $\mathcal{U}_K$ is a family of independent, uniformly on $[0,1]$ distributed r.v. The independence follows from the fact that for the insertion of a key stored in node $\sigma \in D^{(K)}$ only the components less that the *active* component are used. Let $Y = (Y_1, \ldots, Y_s)$ be the vector of the specified components of the query. Corresponding to (1.8) define for $\sigma \in D^{(K)}$

$$\mathbf{1}_\sigma(Y, \mathcal{U}_K) := \prod_{\substack{1 \leq j \leq s \\ \sigma_{r_j}=0}} \mathbf{1}_{\{Y_j < U_{\sigma|(r_j-1)}\}} \prod_{\substack{1 \leq j \leq s \\ \sigma_{r_j}=1}} \mathbf{1}_{\{Y_j \geq U_{\sigma|(r_j-1)}\}}. \tag{1.11}$$

This indicator is one if and only if the subtree $\sigma \in D_K$ has to be inspected for the subsequent search given the query Y and the discriminators $\mathcal{U}_K$. Denote by $(I_\sigma^{(n)})_{\sigma \in D_K}$ the cardinalities of the subtrees on level K if n keys are in the tree. If the levels $0, \dots, K-1$ are not full we arrange $I_\sigma^{(n)} := 0$ for the 'not existing' subtrees $\sigma \in D_K$. Conditionally given full levels and discriminators $\mathcal{U}_K$ the vector $(I_\sigma^{(n)})_{\sigma \in D_K}$ is multinomial distributed. The parameters are given as the probabilities to follow the paths to a certain subtree :

$$\langle \mathcal{U}_K \rangle_\sigma := \prod_{\substack{1 \le j \le K \\ \sigma_j = 0}} U_{\sigma|(j-1)} \prod_{\substack{1 \le j \le K \\ \sigma_j = 1}} (1 - U_{\sigma|(j-1)}), \quad \sigma \in D_K. \tag{1.12}$$

This implies

$$\mathbb{P}^{I^{(n)}} = p_n M(n - 2^K + 1, \langle \mathcal{U}_K \rangle) + (1 - p_n) \mu_n \tag{1.13}$$

with an additional probability measure μ_n on $\mathbb{R}$. A weak law of large numbers follows:

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle \mathcal{U}_K \rangle = (\langle \mathcal{U}_K \rangle_\sigma)_{\sigma \in D_K}. \tag{1.14}$$

The cost of a partial match query satisfies the distributional recursive equation

$$C_n \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) C^{(\sigma)}_{I_\sigma^{(n)}} + N_n. \tag{1.15}$$

Here $Y_1, \dots, Y_s, (\mathcal{U}_K, I^{(n)}), (C_i^{(\sigma)})_{i \in \mathbb{N}}, \sigma \in D_K$ are independent, Y_j, U_σ uniformly distributed on $[0,1]$ $(1 \le j \le s, \sigma \in D^{(K)}), (C_i^{(\sigma)}) \sim (C_i)$ for $\sigma \in D_K$, $I^{(n)}$ as in (1.13), and N_n is the number of nodes traversed during the query on the levels $0, \dots, K-1$, in particular $0 \le N_n \le 2^K$. Define $C_0 := 0$. The mean of the cost C_n has been studied in Flajolet and Puech (1986):

$$\mathbb{E}\, C_n \sim \gamma_u n^{\alpha - 1}, \tag{1.16}$$

with α being the unique solution in $(1,2)$ of the *indicial equation*

$$(\alpha + 1)^s \alpha^{K-s} = 2^K \tag{1.17}$$

and a constant $\gamma_u > 0$ depending on the specification pattern u. γ_u can be approximated numerically (see Flajolet and Puech (1986)). The variance of C_n has been investigated in a more general situation (see next section) in Cunto, Lau, and Flajolet (1989). For the scaling assume

$$\mathrm{Var}(C_n) \sim \beta_u n^{2\alpha - 2} \tag{1.18}$$

with a constant $\beta_u > 0$. This asymptotic is proved in Corollary 1.1.3 with an explicit formula for β_u in terms of α and γ_u. Therefore we introduce

$$X_n := \frac{C_n - \mathbb{E}\, C_n}{n^{\alpha - 1}}. \tag{1.19}$$

A straightforward calculation using (1.15) yields to the distributional recursion for X_n:

$$X_n \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K) \left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} + \gamma_u\right) - \gamma_u + o(1). \qquad (1.20)$$

In (1.20) the (in-)dependencies and distributions are analogously to (1.15). The $o(1)$ depends on randomness but the convergence is uniform. This modified recursion and the convergence of the occurring prefactors (cf. (1.14)) suggest that a limit X of X_n is a solution of the *limiting equation*

$$X \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K) \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1} \left(X^{(\sigma)} + \gamma_u\right) - \gamma_u. \qquad (1.21)$$

Here Y and $\mathcal{U}_K$ are as in (1.15) and $\{X^{(\sigma)}, \sigma \in D_K\}$ is a family of independent r.v. identically distributed as X and independent of $Y, \mathcal{U}_K$. The limiting equation (1.21) allows to define the limiting operator acting on probability measures. This operator has a unique fixed point in a suitably chosen domain. Then convergence of X_n to this fixed point can be established. Let

$$T_u : M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1)$$

$$T_u(\mu) \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K) \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1} \left(Z^{(\sigma)} + \gamma_u\right) - \gamma_u. \qquad (1.22)$$

In (1.22) $Y, \mathcal{U}_K$ and $Z^{(\sigma)}$ are independent, $Z^{(\sigma)} \sim \mu$ for $\sigma \in D_K$ and $Y, \mathcal{U}_K$ are as in (1.21) .

Lemma 1.1.1 *$T_u : M_{0,2} \to M_{0,2}$, with T_u given in (1.22) is a contraction w.r.t. ℓ_2:*

$$\ell_2(T_u(\mu), T_u(\nu)) \leq \xi_u \, \ell_2(\mu, \nu) \quad \textit{for all} \quad \mu, \nu \in M_{0,2}, \qquad (1.23)$$

$$\xi_u = \left(\alpha^s (\alpha - 1/2)^{K-s}\right)^{-1/2} < 1. \qquad (1.24)$$

Proof: This is the special case $t = 0$ of Lemma 1.2.1 in the next section. ■

By Banach's fixed point theorem T_u has a unique fixed point ρ in $M_{0,2}$ and

$$\ell_2(T_u^n(\mu), \rho) \to 0 \qquad (1.25)$$

exponentially fast for any $\mu \in M_{0,2}$. A random variable X with distribution ρ is also called a fixed point of T (compare equation (1.21)).

The proof of the following limit theorem is a typical application of the contraction method. It is given here in detail and referred to it in the next sections for limit laws for variants of the K-d tree. For a unifying setting of related divide and conquer algorithms see section 3 in Rösler (1999).

Theorem 1.1.2 (Limit Theorem for Partial Match Query in random K-d trees) *The normalized number of nodes X_n traversed during a partial match query with specification pattern $u \in \{S, *\}^K$ in a random K-d tree converges w.r.t. ℓ_2 to the unique fixed point X in $M_{0,2}$ of the limiting operator T_u, i.e.*

$$\ell_2(X_n, X) \to 0. \tag{1.26}$$

Proof: Let $X_n^{(\sigma)} \sim X_n, X^{(\sigma)} \sim X$ for $\sigma \in D_K$ such that $(X_n^{(\sigma)}, X^{(\sigma)})$ are optimal couplings of X_n, X, i.e. $\ell_2^2(X_n, X) = \mathbb{E}\,(X_n^{(\sigma)} - X^{(\sigma)})^2$. Let $\mathcal{U}_K := \{U_\sigma : \sigma \in D^{(K)}\}, Y = (Y_1, \dots, Y_s)$ be a family respectively vector of independent r.v. uniformly distributed on $[0,1]$. Furthermore let $I^{(n)}$ be distributed as in (1.13), in particular $I^{(n)}/n \to \langle \mathcal{U}_K \rangle$ in probability as given in (1.14). Finally assume that $(I^{(n)}, \mathcal{U}_K), Y, ((X_n^{(\sigma)}), X^{(\sigma)})(\sigma \in D_K)$ are independent. In order to derive a reduction inequality for $\ell_2(X_n, X)$ we use the L_2-distances of the special representations of X_n and X given by (1.20) resp. (1.21). Then using the independence properties and $\mathbb{E}\,X^{(\sigma)} = \mathbb{E}\,X_n^{(\sigma)} = 0$ we obtain

$$\begin{aligned} \ell_2^2(X_n, X) &\leq \mathbb{E}\left(\sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) \left(\left(\frac{I_\sigma^{(n)}}{n} \right)^{\alpha-1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} + \gamma_u \right) \right.\right. \\ &\qquad \left.\left. - \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1} \left(X^{(\sigma)} + \gamma_u \right) \right) + o(1) \right)^2 \\ &= \mathbb{E} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) \left(\left(\frac{I_\sigma^{(n)}}{n} \right)^{\alpha-1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} + \gamma_u \right) \right. \\ &\qquad \left. - \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1} \left(X^{(\sigma)} + \gamma_u \right) \right)^2 + o(1). \end{aligned} \tag{1.27}$$

The mixed terms are $o(1)$ by independence and $\mathbb{E}\,[(I^{(n)}/n)^{\alpha-1} - \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1}] = o(1)$ for $\sigma \in D_K$. The summands in (1.27) are identically distributed. With a fixed $\sigma \in D_K$ this yields

$$\begin{aligned} &\ell_2^2(X_n, X) \\ &\leq 2^K\, \mathbb{E}\left[\mathbf{1}_\sigma(Y, \mathcal{U}_K) \left(\left(\frac{I_\sigma^{(n)}}{n} \right)^{\alpha-1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} + \gamma_u \right) \right.\right. \\ &\qquad \left.\left. - \langle \mathcal{U}_K \rangle_\sigma^{\alpha-1} \left(X^{(\sigma)} + \gamma_u \right) \right)^2 \right] + o(1) \\ &= 2^K\, \mathbb{E}\left[\mathbf{1}_\sigma(Y, \mathcal{U}_K) \left(\left(\frac{I_\sigma^{(n)}}{n} \right)^{\alpha-1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)} \right) \right.\right. \end{aligned}$$

$$+\left(\left(\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} - \langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\right)\left(X^{(\sigma)}+\gamma_u\right)\right)^2\Bigg] + o(1)$$

$$\begin{aligned} &= 2^K\,\mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{I_\sigma^{(n)}}{n}\right)^{2\alpha-2}\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right)^2\right] \\ &\quad + 2^K\,\mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} - \langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\right)^2\left(X^{(\sigma)}+\gamma_u\right)^2\right] \\ &\quad + 2^K 2\,\mathbb{E}\Bigg[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1}\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right) \\ &\qquad\qquad \times\left(\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} - \langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\right)\left(X^{(\sigma)}+\gamma_u\right)\Bigg] + o(1). \end{aligned} \tag{1.28}$$

With (1.14) it follows

$$\mathbb{E}\left(\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} - \langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\right)^2 \to 0 \;\text{ for }\; n\to\infty. \tag{1.29}$$

Therefore the second summand in (1.28) converges to 0. With the Cauchy-Schwarz inequality and (1.29) the third term in its absolute value is estimated from above by

$$\begin{aligned} &2^K 2\,\mathbb{E}\left[\left(\left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha-1} - \langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\right)^2\left(X^{(\sigma)}+\gamma_u\right)^2\right]^{1/2}\mathbb{E}\left[\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right)^2\right]^{1/2} \\ &= o(1)\,\mathbb{E}\left[\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right)^2\right]^{1/2} \le o(1)\,\mathbb{E}\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right)^2 + o(1). \end{aligned} \tag{1.30}$$

The last inequality holds since both sides are $o(1)$ if the expectation is less than 1. Therefore from (1.28) we derive with $a_n := \ell_2^2(X_n, X)$ and fixed $\sigma \in D_K$

$$\begin{aligned} a_n &\le 2^K\,\mathbb{E}\left[\left(\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{I_\sigma^{(n)}}{n}\right)^{2\alpha-2} + o(1)\right)\left(X^{(\sigma)}_{I_\sigma^{(n)}} - X^{(\sigma)}\right)^2\right] + o(1) \\ &= 2^K\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\mathbf{1}_{\{I_\sigma^{(n)}=i\}}\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right)\left(X_i^{(\sigma)} - X^{(\sigma)}\right)^2\right] + o(1) \end{aligned}$$

$$= \quad 2^K \sum_{i=0}^{n-1} \mathbb{E}\left[\mathbf{1}_{\{I_\sigma^{(n)}=i\}} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] a_i + o(1). \tag{1.31}$$

By (1.14) and an explicit calculation (cf. (1.70), (1.71) for $t = 0$ below) it follows

$$\mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{I_\sigma^{(n)}}{n}\right)^{2\alpha-2}\right] \longrightarrow \mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle\mathcal{U}_K\rangle_\sigma^{2\alpha-2}\right] = \frac{\xi_u^2}{2^K} \tag{1.32}$$

with ξ_u given in (1.24). This implies

$$\begin{aligned}
a_n &\leq 2^K \sum_{i=0}^{n-1} \mathbb{E}\left[\mathbf{1}_{\{I_\sigma^{(n)}=i\}} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] \sup_{1\leq i\leq n-1} a_i + o(1) \\
&= 2^K\, \mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle\mathcal{U}_K\rangle_\sigma^{2\alpha-2}\right] \sup_{1\leq i\leq n-1} a_i + o(1) \\
&= (\xi_u^2 + o(1)) \sup_{1\leq i\leq n-1} a_i + o(1).
\end{aligned} \tag{1.33}$$

Thus $(a_n)_{n\in\mathbb{N}}$ is bounded. Denote $a := \limsup_{n\to\infty} a_n$. Now we can conclude as in Rösler (1991). For a given $\varepsilon > 0$ there exists a $n_0 \in \mathbb{N}$ and $\xi^+ < 1$ with $a_n \leq a + \varepsilon$ and $\xi_u^2 + o(1) \leq \xi^+ < 1$ for all $n \geq n_0$. Then from (1.31) it follows

$$\begin{aligned}
a_n &\leq 2^K \sum_{i=0}^{n_0-1} \mathbb{E}\left[\mathbf{1}_{\{I_\sigma^{(n)}=i\}} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] a_i \\
&\quad + 2^K \sum_{i=n_0}^{n-1} \mathbb{E}\left[\mathbf{1}_{\{I_\sigma^{(n)}=i\}} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] (a+\varepsilon) + o(1) \\
&\leq \xi^+(a+\varepsilon) + o(1).
\end{aligned} \tag{1.34}$$

Now $n \to \infty$ yields $a \leq \xi^+(a + \varepsilon)$, which implies $a = 0$. ■

Convergence in ℓ_2 implies convergence of the second moments. Thus a first order asymptotic for the variance of C_n follows.

Corollary 1.1.3 *The variance of the limiting distribution for the normalized number of nodes traversed during a partial match query with specification pattern $u \in \{S, *\}^K$ in a random K-d tree is given by*

$$\beta_u := \left[\frac{(2\alpha-1)B(\alpha,\alpha)}{\alpha^s(\alpha-1/2)^{K-s} - 1} \sum_{l\in\mathbf{U}} \left(\frac{2(2\alpha-1)}{\alpha^2}\right)^{K-l} \eta_\alpha^{s-s_l} - 1\right]\gamma_u^2 \tag{1.35}$$

with

$$\eta_\alpha = \frac{\alpha\,(8\alpha^2 - 2\alpha - 2 - \alpha(\alpha+1)B(\alpha,\alpha))}{2(\alpha+1)(2\alpha-1)(2\alpha+1)}. \tag{1.36}$$

In (1.35) $\mathbf{U} \subset \{1,\ldots,K\}$ *denotes the set of unspecified components of* u *and* s_l *the number of specified components less than* $l \in \mathbf{U}$. α *and* γ_u *are given by (1.16), (1.17),* $B(\,\cdot\,,\,\cdot\,)$ *denotes the Eulerian beta integral. The variance of the (unscaled) cost* C_n *satisfies*

$$Var(C_n) \sim \beta_u n^{2\alpha-2}. \tag{1.37}$$

Proof: The translation $\tilde{X} := X + \gamma_u$ of the fixed point X of T_u is determined as the unique solution in $M_{\gamma_u,2}$ of the distributional equation

$$\tilde{X} \stackrel{\mathcal{D}}{=} \sum_{\sigma\in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\tilde{X}^{(\sigma)} \tag{1.38}$$

where the independencies and distributions are as in (1.21) and $\tilde{X}^{(\sigma)} \sim \tilde{X}$ for $\sigma \in D_K$. It is

$$\mathrm{Var}(X) = \mathrm{Var}(\tilde{X}) = \mathbb{E}\,\tilde{X}^2 - \gamma_u^2 \tag{1.39}$$

and

$$\begin{aligned}
\mathbb{E}\,\tilde{X}^2 &= \mathbb{E}\left[\sum_{\sigma,\tau\in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\mathbf{1}_\tau(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\langle \mathcal{U}_K\rangle_\tau^{\alpha-1}\tilde{X}^{(\sigma)}\tilde{X}^{(\tau)}\right] \\
&= \mathbb{E}\left[\sum_{\sigma=\tau} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{2\alpha-2}\tilde{X}^2\right] \\
&\quad + \mathbb{E}\left[\sum_{\sigma\neq\tau} \mathbf{1}_\sigma(Y,\mathcal{U}_K)\mathbf{1}_\tau(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha-1}\langle \mathcal{U}_K\rangle_\tau^{\alpha-1}\tilde{X}^{(\sigma)}\tilde{X}^{(\tau)}\right].
\end{aligned} \tag{1.40}$$

Since the summands with $\sigma = \tau$ are identically distributed and with a calculation as in (1.32) the first summand in (1.40) is equal to

$$2^K 2^{-K}\frac{1}{\alpha^s(\alpha-1/2)^{K-s}}\,\mathbb{E}\,\tilde{X}^2. \tag{1.41}$$

The second summand is (cf. (1.11), (1.12))

$$\sum_{\substack{\sigma,\tau\in D_K\\ \sigma\neq\tau}} \mathbb{E}\Bigg[\prod_{\substack{1\le j\le s\\ \sigma r_j=0}} \mathbf{1}_{\{Y_j<U_{\sigma|(r_j-1)}\}} \prod_{\substack{1\le j\le s\\ \sigma r_j=1}} \mathbf{1}_{\{Y_j\ge U_{\sigma|(r_j-1)}\}}$$

$$\prod_{\substack{1\le j\le s\\ \tau r_j=0}} \mathbf{1}_{\{Y_j<U_{\tau|(r_j-1)}\}} \prod_{\substack{1\le j\le s\\ \tau r_j=1}} \mathbf{1}_{\{Y_j\ge U_{\tau|(r_j-1)}\}}$$

$$\prod_{\substack{1\le j\le K\\ \sigma_j=0}} U_{\sigma|(j-1)}^{\alpha-1} \prod_{\substack{1\le j\le K\\ \sigma_j=1}} (1-U_{\sigma|(j-1)})^{\alpha-1}$$

$$\prod_{\substack{1\le j\le K\\ \tau_j=0}} U_{\tau|(j-1)}^{\alpha-1} \prod_{\substack{1\le j\le K\\ \tau_j=1}} (1-U_{\tau|(j-1)})^{\alpha-1}\Bigg]. \tag{1.42}$$

For $\sigma, \tau \in D_K$ with $\sigma \neq \tau$ denote by

$$\begin{aligned} l_{\sigma,\tau} &:= \max\{1 \le j \le K : \sigma|(j-1) = \tau|(j-1)\} \\ &= \min\{1 \le j \le K : \sigma_j \neq \tau_j\} \end{aligned} \tag{1.43}$$

the first component where the vectors σ, τ differ. If $l_{\sigma,\tau}$ is a specified component then the expectation in (1.42) for these σ, τ is zero by disjoint indicator sets. Further denote

$$s_l := \mathrm{card}\{1 \le j \le s : r_j < l\} \tag{1.44}$$

the number of specified components less than l and $\mathbf{U} := \{1, \dots, K\}\backslash\{r_1, \dots, r_s\}$ the set of unspecified components. The distribution of the summand in (1.42) depends only on $l_{\sigma,\tau}$ and the number of specified components greater than $l_{\sigma,\tau}$ in which σ and τ differ. For this number we write

$$h_{\sigma,\tau} := \mathrm{card}\{s_{l_{\sigma,\tau}} + 1 \le j \le s : \sigma_{r_j} \neq \tau_{r_j}\}. \tag{1.45}$$

For given $l \in \{1, \dots, K\}$ and $h \in \{0 \dots, s - s_l\}$ there exist

$$2^K \binom{s-s_l}{h} 2^{K-l-(s-s_l)} \tag{1.46}$$

pairs $(\sigma, \tau) \in D_K \times D_K$ with $\sigma \neq \tau, l_{\sigma,\tau} = l$ and $h_{\sigma,\tau} = h$. For these pairs the summands in (1.42) are identically distributed. With Y, U, V independent and uniformly distributed on $[0, 1]$ these expectations are given by

$$\begin{aligned} & \mathbb{E}\left[\mathbf{1}_{\{Y<U\}}U^{2\alpha-2}\right]^{s_l} \mathbb{E}\left[U^{2\alpha-2}\right]^{l-1-s_l} \mathbb{E}\left[U^{\alpha-1}(1-U)^{\alpha-1}\right] \\ \times\ & \mathbb{E}\left[U^{\alpha-1}V^{\alpha-1}\right]^{K-l-(s-s_l)} \mathbb{E}\left[\mathbf{1}_{\{Y<U\}}U^{\alpha-1}\mathbf{1}_{\{Y\ge V\}}(1-V)^{\alpha-1}\right]^h \\ \times\ & \mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\mathbf{1}_{\{Y<V\}}U^{\alpha-1}V^{\alpha-1}\right]^{s-s_l-h}. \end{aligned} \tag{1.47}$$

Explicit calculations yield

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}U^{2\alpha-2}\right] = 1/(2\alpha), \quad \mathbb{E}\left[U^{2\alpha-2}\right] = 1/(2\alpha-1),\\
\mathbb{E}\left[U^{\alpha-1}(1-U)^{\alpha-1}\right] = B(\alpha,\alpha), \quad \mathbb{E}\left[U^{\alpha-1}V^{\alpha-1}\right] = 1/(\alpha^2),\\
\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}U^{\alpha-1}\mathbf{1}_{\{Y\geq V\}}(1-V)^{\alpha-1}\right] = \frac{\alpha-1+\alpha B(\alpha,\alpha+2)}{\alpha^2(\alpha+1)},\\
\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\mathbf{1}_{\{Y<V\}}U^{\alpha-1}V^{\alpha-1}\right] = 2/((\alpha+1)(2\alpha+1)). \qquad (1.48)
\end{aligned}$$

Altogether for the second summand in (1.40) we derive

$$\sum_{l\in\mathbf{U}} 2^{2K-l-(s-s_l)}B(\alpha,\alpha)\left(\frac{1}{\alpha^2}\right)^{K-l-(s-s_l)}\left(\frac{1}{2\alpha-1}\right)^{l-1-s_l} \qquad (1.49)$$

$$\times\sum_{h=0}^{s-s_l}\binom{s-s_l}{h}\left(\frac{\alpha-1+\alpha B(\alpha,\alpha+2)}{\alpha(\alpha+1)}\right)^h\left(\frac{2}{(\alpha+1)(2\alpha+1)}\right)^{s-s_l-h}.$$

Using the binomial formula, $B(\alpha,\alpha+2) = (\alpha+1)/(2(2\alpha+1))B(\alpha,\alpha)$ and some simplifications this is

$$\sum_{l\in\mathbf{U}} 2^{2K-l-s}B(\alpha,\alpha)\left(\frac{1}{\alpha}\right)^{2(K-l)+s_l}\left(\frac{1}{2\alpha-1}\right)^{l-1-s_l}$$

$$\times\left(\frac{8\alpha^2-2\alpha-2-\alpha(\alpha+1)B(\alpha,\alpha)}{2(\alpha+1)(2\alpha-1)(2\alpha+1)}\right)^{s-s_l}. \qquad (1.50)$$

With (1.39)–(1.41) this leads to the stated variance β_u. By convergence of the second moments of X_n we conclude

$$\begin{aligned}
\mathrm{Var}(C_n) &= \mathrm{Var}(n^{\alpha-1}X_n) = \mathrm{Var}(X_n)n^{2\alpha-2} = (\mathrm{Var}(X)+o(1))n^{2\alpha-2}\\
&\sim \beta_u n^{2\alpha-2}. \qquad (1.51)
\end{aligned}$$

■

1.2 The locally balanced K-d-t tree

The K-d-t trees introduced in Cunto, Lau and Flajolet (1989) are intermediate structures between the original K-d tree investigated in the previous section and fully balanced K-d trees generated by a total reorganization of the tree (see Bentley (1975)). Such fully balanced K-d trees achive an *optimal* exponent for the expected cost of a partial match query:

$$\mathbb{E}\,C_n \sim \text{const}\ n^{1-s/K} \qquad (1.52)$$

where again $1 \leq s \leq K-1$ is the number of specified components of an arbitrary specification pattern. For the classical K-d tree the asymptotic of Flajolet and Puech (1986) for the average cost (cf. (1.16)) can be written in the form

$$\mathbb{E}\, C_n \sim \gamma_u n^{1-s/K+\theta(s/K)} \tag{1.53}$$

with a strictly positive function $\theta(x)$ for $0 < x < 1$ with maximum value ≤ 0.07. The K-d-t trees have been introduced to improve this performance through a local online reorganization method, which only adds a constant extra effort to updating operations. A K-d-t tree is a K-d tree wherein each subtree of size greater than $2t$ has at least t nodes on each side. In particular, if a subtree has $2t+1$ nodes then the key with the median of the components which are used as discriminator at this level is stored in its root. For $t = 0$ the K-d-t tree coincides with the K-d tree. For details of an implementation of such a structure we refer to Cunto, Lau, and Flajolet (1989). As a main result Cunto et al. derive

$$\mathbb{E}\, C_n^{(t)} \sim \gamma_{t,u} n^{1-s/K+\theta(s/K,t)} \tag{1.54}$$

where $C_n^{(t)}$ denotes the cost of a partial match query (depending on the specification pattern u) in a K-d-t tree with n nodes inserted. $\gamma_{t,u}$ is a positive constant and $\theta(x,t)$ a positive function for $x \in (0,1), t \in \mathbb{N}$ with

$$\theta(x,t) \to 0 \quad \text{for} \quad t \to \infty. \tag{1.55}$$

From this point of view the K-d-t trees improve the K-d tree up to the behavior of a fully balanced K-d tree for $t \to \infty$.

The analysis of a partial match query in random K-d trees given in the previous section can be generalized to random K-d-t trees. Again inspect the levels $0, \dots, K-1$ for a search followed by a recursive search in some of the subtrees on level K, where the original specification pattern can be used. Assume a fixed specification pattern $u \in \{S, *\}$ and suppress the dependency on u in the notation. The nodes and subtrees are numbered as in (1.7). Analogously to (1.9) let $p_n^{(t)}$ denote the probability that in a random K-d-t tree with n keys inserted all subtrees on level $K-1$ have more than $2t$ nodes. By the local reorganization this means that all subtrees on level K have at least t nodes. Then $p_n^{(t)} \to 1$ for $n \to \infty$. In this case the discriminators on the levels $0, \dots, K-1$ occur as the median of $2t+1$ independent uniformly on $[0,1]$ distributed r.v. Therefore they are beta$(t+1, t+1)$ distributed and independent for the same reason as in the case $t = 0$ in the previous section. Thus denote by $\mathcal{U}_K := \{U_\sigma : \sigma \in D^{(K)}\}$ a family of independent, beta$(t+1, t+1)$ distributed r.v. which serves as the discriminators on the levels $0, \dots, K-1$. Then the cardinalities of the subtrees $(I_\sigma^{(n)})_{\sigma \in D_K}$ on level K, n keys being inserted in the tree, satisfy

$$\mathbb{P}^{I^{(n)}} = p_n^{(t)} M(n - 2^K + 1, \langle \mathcal{U}_K \rangle) + (1 - p_n^{(t)})\mu_n \tag{1.56}$$

with a probability measure $\mu_n^{(t)}$. With the notation (1.12) it follows

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle \mathcal{U}_K \rangle = (\langle \mathcal{U}_K \rangle_\sigma)_{\sigma \in D_K}. \tag{1.57}$$

Denote the specified components in the query by $Y = (Y_1, \dots, Y_s)$. Then $Y_1, \dots, Y_s$ are independent and uniformly distributed on $[0,1]$. The cost $C_n^{(t)}$ of a partial match query then satisfies using notation (1.11)

$$C_n^{(t)} \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) C^{(t,\sigma)}_{I_\sigma^{(n)}} + N_n^{(t)}. \tag{1.58}$$

Here $Y, (\mathcal{U}_K, I^{(n)}), (C_i^{(t,\sigma)})_{i \in \mathbb{N}}, \sigma \in D_K$ are independent, $(C_i^{(t,\sigma)}) \sim (C_i^{(t)})$ for $\sigma \in D_K$, $I^{(n)}$ is as in (1.56) and $N_n^{(t)}$ is the number of nodes traversed during the query on the levels $0, \dots, K-1$, in particular $0 \le N_n^{(t)} \le 2^K$. The asymptotic for the mean of $C_n^{(t)}$ (cf. (1.54)) given in Cunto et al. (1989) is

$$\mathbb{E}\, C_n^{(t)} \sim \gamma_{t,u} n^{\alpha_t - 1} \tag{1.59}$$

where $\gamma_{t,u} > 0$ and $\alpha_t \in (1,2)$ is determined by the indicial equation

$$\left((\alpha_t + t)^{\overline{t+1}}\right)^{K-s} \left((\alpha_t + t + 1)^{\overline{t+1}}\right)^{s} = \left((t+2)^{\overline{t+1}}\right)^{K}. \tag{1.60}$$

Here $x^{\overline{t}}$ denotes the upper factorial power, i.e. $x^{\overline{t}} := x(x+1) \cdot \ldots \cdot (x+t-1)$ for $x \in \mathbb{R}$ and $t \in \mathbb{N}$. Observe that we write $\alpha_t - 1$ for the exponent in (1.59) instead of α as in Cunto et al. As an asymptotic of the variance Cunto et al. derived

$$\mathrm{Var}(C_n^{(t)}) = C_{2,t} n^{2\alpha_t - 2} + C_{1,t} n^{\alpha_t - 1}(1 + o(1)) \tag{1.61}$$

with $C_{1,t} > 0$ and $C_{2,t} \to 0$ for $t \to \infty$. The normalized cost (assume a fixed t)

$$X_n := \frac{C_n^{(t)} - \mathbb{E}\, C_n^{(t)}}{n^{\alpha_t - 1}} \tag{1.62}$$

satisfies the distributional recursive equation

$$X_n \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) \left(\frac{I_\sigma^{(n)}}{n}\right)^{\alpha_t - 1} \left(X^{(\sigma)}_{I_\sigma^{(n)}} + \gamma_{t,u}\right) - \gamma_{t,u} + o(1). \tag{1.63}$$

with (in-)dependencies and distributions as in (1.58). By the convergence of $I^{(n)}/n$ in (1.57) we are led to the limiting operator

$$T_{t,u} : M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1)$$

$$T_u(\mu) \stackrel{\mathcal{D}}{=} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y, \mathcal{U}_K) \langle \mathcal{U}_K \rangle_\sigma^{\alpha_t - 1} \left(Z^{(\sigma)} + \gamma_{t,u}\right) - \gamma_{t,u}. \tag{1.64}$$

Here $Y_1, \dots, Y_s, U_\sigma(\sigma \in D^{(K)}), Z^{(\sigma)}(\sigma \in D_K)$ are independent, Y_i uniformly distributed on $[0,1]$, $U_\sigma \sim \mathrm{beta}(t+1, t+1)$ and $Z^{(\sigma)} \sim \mu$.

Lemma 1.2.1 *$T_{t,u} : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. ℓ_2:*

$$\ell_2(T_{t,u}(\mu), T_{t,u}(\nu)) \le \xi_{t,u}\, \ell_2(\mu,\nu) \quad \textit{for all} \quad \mu,\nu \in M_{0,2}, \tag{1.65}$$

$$\xi_{t,u} = \left(\frac{\left((t+2)^{\overline{t+1}}\right)^K}{\left((2\alpha_t+t)^{\overline{t+1}}\right)^s \left((2\alpha_t+t-1)^{\overline{t+1}}\right)^{K-s}} \right)^{1/2} < 1. \tag{1.66}$$

Proof: First we proof that $T_{t,u} : M_{0,2} \to M_{0,2}$ is well defined. Obviously Var $T(\mu) < \infty$. The summands in (1.64) are identically distributed. This implies

$$\mathbb{E}\, T_{t,u}(\mu) \;=\; 2^K\, \mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha_t-1}\right]\gamma_{t,u} - \gamma_{t,u}. \tag{1.67}$$

It is

$$\mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha_t-1}\right] = \mathbb{E}\left[\mathbf{1}_{\{Y_1<U_\emptyset\}} U_\emptyset^{\alpha_t-1}\right]^s \mathbb{E}\left[U_\emptyset^{\alpha_t-1}\right]^{K-s}. \tag{1.68}$$

Explicit calculations yield

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{Y_1<U_\emptyset\}} U_\emptyset^{\alpha_t-1}\right] &= \frac{B(\alpha_t+t+1,t+1)}{B(t+1,t+1)}\\ &= \left((\alpha_t+t+1)^{\overline{t+1}}\right)^{-1}(t+1)^{\overline{t+1}},\end{aligned}$$

$$\mathbb{E}\left[U_\emptyset^{\alpha_t-1}\right] = \frac{B(\alpha_t+t,t+1)}{B(t+1,t+1)} = \left((\alpha_t+t)^{\overline{t+1}}\right)^{-1}(t+1)^{\overline{t+1}}. \tag{1.69}$$

Therefore

$$\begin{aligned}&2^K\,\mathbb{E}\left[\mathbf{1}_\sigma(Y,\mathcal{U}_K)\langle \mathcal{U}_K\rangle_\sigma^{\alpha_t-1}\right]\\ &= 2^K\left((t+1)^{\overline{t+1}}\right)^K\left((\alpha_t+t+1)^{\overline{t+1}}\right)^{-s}\left((\alpha_t+t)^{\overline{t+1}}\right)^{-(K-s)}\\ &= \left((t+2)^{\overline{t+1}}\right)^K\left((\alpha_t+t+1)^{\overline{t+1}}\right)^{-s}\left((\alpha_t+t)^{\overline{t+1}}\right)^{-(K-s)}\\ &= 1\end{aligned} \tag{1.70}$$

where the indicial equation (1.60) has been used. This implies $\mathbb{E}\, T_{t,u}(\mu) = 0$, so $T_{t,u} : M_{0,2} \to M_{0,2}$ is well defined.

In order to prove contractivity let $\mu,\nu \in M_{0,2}$ and let $(W^{(\sigma)},\, Z^{(\sigma)})$, Y, $\mathcal{U}_K$ be independent, $(W^{(\sigma)},\, Z^{(\sigma)})$ be optimal ℓ_2–couplings of (μ,ν), i.e. $W^{(\sigma)} \sim \mu$,

$Z^{(\sigma)} \sim \nu$ and $\ell_2^2(\mu,\nu) = \mathbb{E}\,(W^{(\sigma)} - Z^{(\sigma)})^2$ for $\sigma \in D_K$, Y and $\mathcal{U}_K$ as in (1.64). Then using the independence properties and $\mathbb{E}\,W^{(\sigma)} = \mathbb{E}\,Z^{(\sigma)} = 0$ we conclude

$$\begin{aligned}
&\ell_2^2(T_{t,u}(\mu), T_{t,u}(\nu)) \\
&= \mathbb{E} \sum_{\sigma \in D_K} \mathbf{1}_\sigma(Y,\mathcal{U}_K) \langle \mathcal{U}_K \rangle_\sigma^{2\alpha_t - 2} (W^{(\sigma)} - Z^{(\sigma)})^2 \\
&= 2^K\, \mathbb{E}\left[\mathbf{1}_{\{Y_1 < U_\emptyset\}} U_\emptyset^{2\alpha_t-2}\right]^s \mathbb{E}\left[U_\emptyset^{2\alpha_t-2}\right]^{K-s} \ell_2^2(\mu,\nu) \\
&= 2^K \left(\frac{(t+1)^{\overline{t+1}}}{(2\alpha_t+t)^{\overline{t+1}}}\right)^s \left(\frac{(t+1)^{\overline{t+1}}}{(2\alpha_t+t-1)^{\overline{t+1}}}\right)^{K-s} \ell_2^2(\mu,\nu) \\
&= \frac{\left((t+2)^{\overline{t+1}}\right)^K}{\left((2\alpha_t+t)^{\overline{t+1}}\right)^s \left((2\alpha_t+t-1)^{\overline{t+1}}\right)^{K-s}}\, \ell_2^2(\mu,\nu). \qquad (1.71)
\end{aligned}$$

This yields assertion (1.65). $\alpha_t > 1$ implies $2\alpha_t + t > \alpha_t + t + 1$ and $2\alpha_t + t - 1 > \alpha_t + t$. Together with the indicial equation (1.60) we deduce

$$\begin{aligned}
\xi_{t,u}^2 &= \left((t+2)^{\overline{t+1}}\right)^K \left((2\alpha_t+t)^{\overline{t+1}}\right)^{-s} \left((2\alpha_t+t-1)^{\overline{t+1}}\right)^{-(K-s)} \\
&< \left((t+2)^{\overline{t+1}}\right)^K \left((\alpha_t+t+1)^{\overline{t+1}}\right)^{-s} \left((\alpha_t+t)^{\overline{t+1}}\right)^{-(K-s)} \\
&= 1. \qquad (1.72)
\end{aligned}$$

■

By Banach's fixed point theorem there exists a unique fixed point X of $T_{t,u}$. Similarly to the proof of Theorem 1.1.2 convergence of $(X_n)_{n\in\mathbb{N}}$ to X in the ℓ_2-metric can be established.

Theorem 1.2.2 (Limit Theorem for Partial Match Query in random K-d-t trees) *The normalized number of nodes X_n traversed during a partial match query with specification pattern $u \in \{S,*\}^K$ in a random K-d-t tree converges w.r.t. ℓ_2 to the unique fixed point X in $M_{0,2}$ of the limiting operator $T_{t,u}$, i.e.*

$$\ell_2(X_n, X) \to 0. \qquad (1.73)$$

In particular this implies that the constants $C_{2,t}$ in (1.61) are strictly positive:

Corollary 1.2.3 *The variance of the cost $C_n^{(t)}$ of a partial match query with specification pattern $u \in \{S,*\}^K$ in a random K-d-t tree satisfies*

$$Var(C_n^{(t)}) \sim \beta_{t,u} n^{2\alpha-2} \qquad (1.74)$$

with a constant $\beta_{t,u} > 0$.

Proof: The constant $\beta_{t,u}$ is derived to be the variance of the fixed point X of the operator $T_{t,u}$ similarly to (1.51). X is nondegenerated, thus $\beta_{t,u} > 0$. ■

In principle it is possible (but very tiresome) to calculate $\beta_{t,u}$ in terms of $\gamma_{t,u}, \alpha$ and t following the proof of Corollary 1.1.3.

1.3 The random relaxed K-d tree

Updating operations in classical K-d trees are very laborious. Especially deleting near to the root causes tedious reorganization operations in the whole tree. Furthermore the distribution of the random tree may be changed executing updating operations. This means that the analysis of certain parameters of the original tree built up by the data does not hold true after an alternation of deletions and new insertions. To overcome these problems Duch, Estivill-Castro, and Martínez (1998) recently introduced the *random relaxed K-d tree*. As for the classical K-d tree at each node a certain attribute of the keys is used as the discriminator for the insertion process and queries. Instead of rotating cyclically the components as for the K-d tree for the relaxed K-d tree (r-K-d tree) a relaxation is introduced. At each node additionally to the key a discriminant $j \in \{1, \ldots, K\}$ is stored. This discriminant indicates which component of the keys is drawn for the splitting process at this node. In the *random* r-K-d tree these discriminants are taken uniformly at random in $\{1, \ldots, K\}$ and independent at each node. For details and (randomized) procedures for building up and updating the random r-K-d tree without changing its distribution see Duch et al. (1998).

The performance of some associated queries in random r-K-d trees has already been investigated in Duch et al. (1998) and Martínez, Panholzer and Prodinger (1998). A partial match query does not depend on the particular specification pattern. It only depends on s/K where $1 \leq s \leq K-1$ is the number of components specified in the query. Let $\varrho := s/K$. For the cost C_n of a partial match query in a random r-K-d tree with n nodes inserted Duch et al. (1998) and Martínez et al. (1998) derive

$$\mathbb{E}\, C_n = \beta n^{\alpha-1} + \mathcal{O}(1) \tag{1.75}$$

where

$$\alpha = \frac{1}{2} + \frac{1}{2}\sqrt{9 - 8\varrho} \tag{1.76}$$

is a solution of the indicial equation

$$\alpha^2 - \alpha - 2(1-\varrho) = 0 \tag{1.77}$$

and

$$\beta = \frac{\Gamma(2\alpha - 1)}{(1-\varrho)\alpha\Gamma^3(\alpha)}. \tag{1.78}$$

Furthermore

$$\mathrm{Var}(C_n) \sim v n^{2\alpha-2} \tag{1.79}$$

and v has an elaborated representation in terms of α (see Theorem 3 in Martínez et al. (1998)). In order to deduce a distributional recursive equation for the cost C_n observe that the random discriminant J at the root of the tree indicates a specified component of the query with probability s/K. In this case the subsequent search considers only one of the subtrees depending on the key comparison at the root. Otherwise the query has to follow both subtrees. Denote by U the Jth component of the key stored in the root and by Y the Jth component of the query if this component is specified. Y, U are independent and uniformly distributed on $[0, 1]$. The subtrees of the random r-K-d tree are given their cardinality again random r-K-d trees. Denote by B a Bernoulli r.v. with success probability s/K, $B \sim B(1, s/K)$. Then the following recursion is valid:

$$C_n \stackrel{\mathcal{D}}{=} B\left(\mathbf{1}_{\{Y<U\}} C^{(1)}_{I_1^{(n)}} + \mathbf{1}_{\{Y\geq U\}} C^{(2)}_{I_2^{(n)}}\right) + (1-B)\left(C^{(1)}_{I_1^{(n)}} + C^{(2)}_{I_2^{(n)}}\right) + 1. \tag{1.80}$$

Here $B, Y, (U, I^{(n)}), (C_i^{(1)}), (C_i^{(2)})$ are independent, $(C_i^{(1)}), (C_i^{(2)}) \sim (C_i)$ and the cardinalities $I^{(n)} = (I_1^{(n)}, I_2^{(n)})$ of the subtrees of the root satisfy

$$\mathbb{P}^{I^{(n)}\,|\,U=u} = M(n-1, u, (1-u)). \tag{1.81}$$

For the normalized cost

$$X_n := \frac{C_n - \mathbb{E}\, C_n}{n^{\alpha-1}} \tag{1.82}$$

it follows

$$\begin{aligned} X_n \;\stackrel{\mathcal{D}}{=}\; & B\Bigg(\mathbf{1}_{\{Y<U\}}\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\Big(X^{(1)}_{I_1^{(n)}} + \beta\Big) + \mathbf{1}_{\{Y\geq U\}}\left(\frac{I_2^{(n)}}{n}\right)^{\alpha-1}\Big(X^{(2)}_{I_2^{(n)}} + \beta)\Big)\Bigg) \\ & + \; (1-B)\Bigg(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}} + \beta\right) + \left(\frac{I_2^{(n)}}{n}\right)^{\alpha-1}\left(X^{(2)}_{I_2^{(n)}} + \beta\right)\Bigg) \\ & - \beta + o(1). \end{aligned} \tag{1.83}$$

The (in-)dependencies and distributions are analogously to (1.80). The $o(1)$ depends on randomness but the convergence is uniformly. By (1.81) for the scaled subtree sizes $I^{(n)}/n$ a weak law of large numbers holds:

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} (U, 1-U). \tag{1.84}$$

Therefore the limiting operator is given by

$$\begin{aligned} T &: M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1) \\ T(\mu) &\stackrel{\mathcal{D}}{=} B\Big(\mathbf{1}_{\{Y<U\}} U^{\alpha-1}\left(Z^{(1)}+\beta\right) + \mathbf{1}_{\{Y \geq U\}}(1-U)^{\alpha-1}\left(Z^{(2)}+\beta\right)\Big) \\ &\quad + (1-B)\Big(U^{\alpha-1}\left(Z^{(1)}+\beta\right) + (1-U)^{\alpha-1}\left(Z^{(2)}+\beta\right)\Big) \\ &\quad - \beta. \end{aligned} \tag{1.85}$$

Here $B, Y, U, Z^{(1)}, Z^{(2)}$ are independent, $B \sim B(1, s/K), Y, U$ are uniformly distributed on $[0,1]$ and $Z^{(1)}, Z^{(2)} \sim \mu$.

Lemma 1.3.1 *$T : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. ℓ_2:*

$$\ell_2(T(\mu), T(\nu)) \leq \xi\, \ell_2(\mu, \nu) \quad \textit{for all} \quad \mu, \nu \in M_{0,2}, \tag{1.86}$$

$$\xi = \left(2\, \frac{1-\varrho+\sqrt{9-8\varrho}}{9-8\varrho+\sqrt{9-8\varrho}}\right)^{1/2} < 1. \tag{1.87}$$

Proof: $T : M_{0,2} \to M_{0,2}$ is well defined: $\mathrm{Var}(T(\mu)) < \infty$ is obvious. It is

$$\begin{aligned} \mathbb{E}\, T(\mu) &= \varrho 2\, \mathbb{E}\, [\mathbf{1}_{\{Y<U\}} U^{\alpha-1}]\beta + (1-\varrho) 2\, \mathbb{E}\, [U^{\alpha-1}]\beta - \beta \\ &= \left[\frac{2\varrho}{\alpha+1} + \frac{2(1-\varrho)}{\alpha}\right]\beta - \beta \\ &= \frac{2\varrho\alpha + 2(1-\varrho)(\alpha+1)}{\alpha^2+\alpha}\beta - \beta \\ &= \frac{2\alpha + 2(1-\varrho)}{2\alpha + 2(1-\varrho)}\beta - \beta. \end{aligned} \tag{1.88}$$

For the last equality the indicial equation (1.77) has been used. Therefore, $T : M_{0,2} \to M_{0,2}$ is well defined. To prove contractivity let $\mu, \nu \in M_{0,2}$ and let $(W^{(1)}, Z^{(1)})$, $(W^{(2)}, Z^{(2)})$, Y, U be independent, Y, U uniformly distributed on $[0,1]$. Let $(W^{(i)}, Z^{(i)})$ be optimal ℓ_2–couplings of (μ, ν), i.e. $W^{(i)} \sim \mu$, $Z^{(i)} \sim \nu$, and $\ell_2^2(\mu,\nu) = \mathbb{E}\,(W^{(i)} - Z^{(i)})^2$ for $i = 1,2$. Then using the independence properties and $\mathbb{E}\,W^{(i)} = \mathbb{E}\,Z^{(i)} = 0$

$$
\begin{aligned}
&\ell_2^2(T(\mu), T(\nu)) \\
&\leq \mathbb{E}\left(B\left(\mathbf{1}_{\{Y<U\}} U^{\alpha-1}(W^{(1)} - Z^{(1)}) + \mathbf{1}_{\{Y\geq U\}} (1-U)^{\alpha-1}(W^{(2)} - Z^{(2)}) \right)\right. \\
&\qquad \left. + (1-B)\left(U^{\alpha-1}(W^{(1)} - Z^{(1)}) + (1-U)^{\alpha-1}(W^{(2)} - Z^{(2)}) \right)\right)^2 \\
&= \mathbb{E}\left[B\mathbf{1}_{\{Y<U\}} U^{2\alpha-2}(W^{(1)} - Z^{(1)})^2 + B\mathbf{1}_{\{Y\geq U\}}(1-U)^{2\alpha-2}(W^{(2)} - Z^{(2)})^2 \right. \\
&\qquad \left. + (1-B)U^{2\alpha-2}(W^{(1)} - Z^{(1)})^2 + (1-B)(1-U)^{2\alpha-2}(W^{(2)} - Z^{(2)})^2 \right] \\
&= 2\left(\mathbb{E}\,[B\mathbf{1}_{\{Y<U\}}U^{2\alpha-2}] + \mathbb{E}\,[(1-B)U^{2\alpha-2}] \right) \ell_2^2(\mu,\nu). \qquad (1.89)
\end{aligned}
$$

It is

$$
\mathbb{E}\,[B\,\mathbf{1}_{\{Y<U\}}U^{2\alpha-2}] = \frac{\varrho}{2\alpha}, \quad \mathbb{E}\,[(1-B)U^{2\alpha-2}] = \frac{1-\varrho}{2\alpha-1}. \qquad (1.90)
$$

This implies

$$
\ell_2^2(T(\mu), T(\nu)) \leq \left(\frac{\varrho}{\alpha} + \frac{1-\varrho}{2\alpha-1} \right) \ell_2^2(\mu,\nu). \qquad (1.91)
$$

The representation (1.76) of α leads to the stated prefactor ξ in (1.87). An explicit calculation shows $\xi < 1$. ■

By Banach's fixed point theorem T has a unique fixed point ϱ in $M_{0,2}$. The representation of the limiting operator can be simplified. We have

$$
\begin{aligned}
T(\mu) &\stackrel{\mathcal{D}}{=} B\,U^{\frac{\alpha-1}{2}}\left(Z^{(1)} + \beta\right) \qquad (1.92) \\
&\qquad + (1-B)\Big(U^{\alpha-1}\left(Z^{(1)} + \beta\right) + (1-U)^{\alpha-1}\left(Z^{(2)} + \beta\right) \Big) - \beta.
\end{aligned}
$$

with $B, U, Z^{(1)}, Z^{(2)}$ being independent, $B \sim B(1, s/K), U$ uniformly distributed on $[0,1]$ and $Z^{(1)}, Z^{(2)} \sim \mu$. The proof follows from an elementary calculation observing that the sets of the indicator functions in (1.85) are disjoint and $\sqrt{U}$ has the density $2x$ for $0 \leq x \leq 1$. By an additional translation it follows that X is a fixed point of T in $M_{0,2}$ if and only if $\widetilde{X} := X + \beta$ is a fixed point of

$$
\widetilde{T}(\mu) \stackrel{\mathcal{D}}{=} BU^{\frac{\alpha-1}{2}}Z^{(1)} + (1-B)\Big(U^{\alpha-1}Z^{(1)} + (1-U)^{\alpha-1}Z^{(2)} \Big) \qquad (1.93)
$$

in $M_{\beta,2}$.

The convergence in ℓ_2 of X_n can be established following the scheme of the proof of Theorem 1.1.2. The corresponding quantities can easily be identified (cf. section 3 in Rösler (1999)).

Theorem 1.3.2 (Limit Theorem for Partial Match Query in random relaxed K-d trees) *The normalized number X_n of nodes traversed during a partial match query in a random r-K-d tree with $1 \leq s \leq K-1$ components specified converges w.r.t. ℓ_2 to the unique fixed point X in $M_{0,2}$ of the limiting operator T, i.e.*

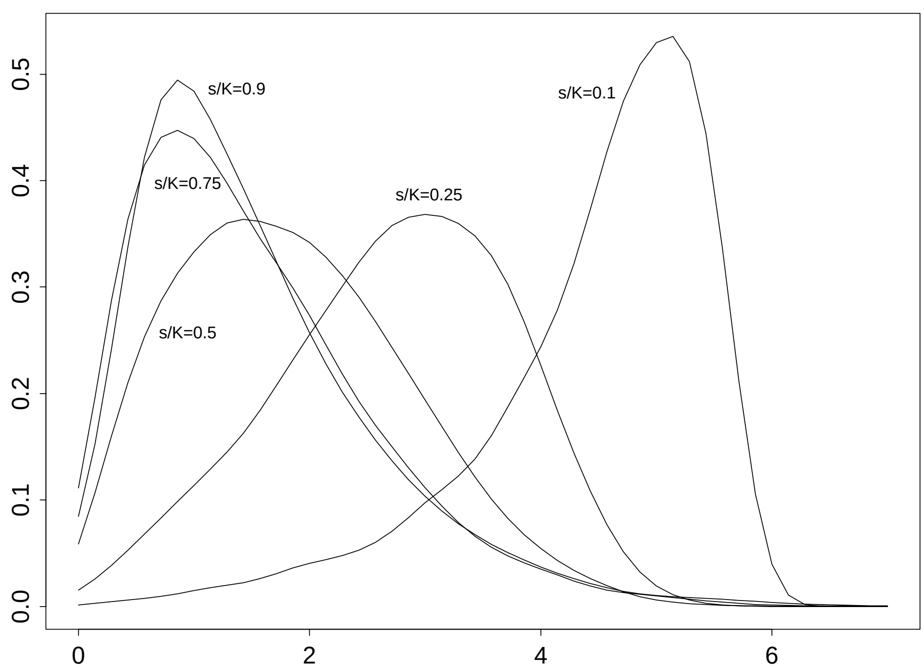
$$\ell_2(X_n, X) \to 0. \tag{1.94}$$

The translated limiting distribution $\widetilde{X} := X + \beta$ is the unique solution in $M_{\beta,2}$ of the limiting equation

$$Z \stackrel{\mathcal{D}}{=} BU^{\frac{\alpha-1}{2}} Z^{(1)} + (1-B)\Big(U^{\alpha-1}Z^{(1)} + (1-U)^{\alpha-1}Z^{(2)}\Big) \tag{1.95}$$

with $B, U, Z^{(1)}, Z^{(2)}$ independent, $B \sim B(1, s/K)$, U uniformly distributed on $[0,1]$, and $Z^{(1)}, Z^{(2)} \sim Z$.

Since the parameters α and β have explicit representations in terms of s/K it is possible to sample the limit distributions approximately for special values of s/K iterating the limiting equation.



The approximations of the densities of the limiting equations for the parameters $\varrho = s/K = 0.1, 0.25, 0.5, 0.75, 0.9$ were produced by iterating 10 times the translated limiting operator $\widetilde{T}$ starting with δ_β, the Dirac measure in β. 10 000 samples of $\widetilde{T}^{10}(\delta_\beta)$ were produced and a standard smoothing-routine of S-Plus on the histogram of the data was applied. The densities have been scaled to have variance 1, i.e. the plotted densities correspond to the limits of $C_n/\sqrt{\mathrm{Var}(C_n)}$.

1.4 The multidimensional Quadtree

Consider a partial match query for a d-dimensional random quadtree with $1 \leq s \leq d-1$ components specified. By symmetry we can assume w.l.g. these are the first s coordinates. Then after a comparison of the search pattern with an internal node of the quadtree we have to inspect 2^{d-s} subtrees at this node for the subsequent search. A node $u \in [0,1]^d$ partitions the quadrant it belongs to into 2^d subquadrants. Let the number of a subquadrant be given by

$$\sum_{i=1}^{d} 2^{d-i} \mathbf{1}_{\{u_i \leq s_i\}}, \qquad u = (u_i), s = (s_i) \tag{1.96}$$

if s is a point in this subquadrant. A key p is inserted in the kth subtree if it belongs to the kth subquadrant. For the binary representation of $0 \leq k \leq 2^d - 1$

$$k = \sum_{i=1}^{d} a_i 2^{d-i}, \qquad a_i = a_i(k) \in \{0,1\} \tag{1.97}$$

let

$$E(k) := \{i \in \{1 \ldots, d\} \mid a_i(k) = 1\}, \tag{1.98}$$

$$N(k) := \{i \in \{1 \ldots, d\} \mid a_i(k) = 0\}. \tag{1.99}$$

Then equivalently, p is inserted in the kth subtree of a node w if $p_i \geq w_i$ for all $i \in E(k)$ and $p_i < w_i$ for all $i \in N(k)$.
For the root $u \in [0,1]^d$ of the tree the volumina of the generated quadrants are given by

$$\langle u \rangle_k := \prod_{i \in N(k)} u_i \prod_{i \in E(k)} (1 - u_i). \tag{1.100}$$

$\langle u \rangle := (\langle u \rangle_0, \ldots, \langle u \rangle_{2^d - 1})$ denotes the vector of the generated volumes. The vector $I^{(n)}$ of the cardinalities of the subtrees of a random d-dimensional quadtree with n nodes is conditionally given the root U multinomial distributed:

$$\mathbb{P}^{I^{(n)}|U=u} = M(n-1, \langle u \rangle), \tag{1.101}$$

where U, the first key to be inserted, is uniformly distributed on $[0,1]^d$.The weak law of large numbers for $I^{(n)}/n$ follows:

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle U \rangle = (\langle U \rangle_0 \ldots, \langle U \rangle_{2^d-1}). \tag{1.102}$$

We denote the s specified components of the search pattern by $Y = (Y_1, \ldots, Y_s)$. The Y_i are independent, uniformly distributed on $[0,1]$ and independent of the random quadtree. In order to give a concise form of the recursive distributional equation for the number C_n of nodes traversed during a partial match query in a random d-dimensional quadtree with n nodes and $1 \le s \le d-1$ components specified we define for $j_1, \ldots, j_s \in \{0,1\}$

$$\mathbf{1}_{j_1,\ldots,j_s}(U,Y) := \prod_{\substack{1\le i\le s\\ j_i=0}} \mathbf{1}_{\{Y_i<U_i\}} \prod_{\substack{1\le i\le s\\ j_i=1}} \mathbf{1}_{\{Y_i\ge U_i\}}. \tag{1.103}$$

Then the cost C_n satisfies

$$C_n \stackrel{\mathcal{D}}{=} \sum_{j_1,\ldots,j_s=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y) \sum_{j_{s+1},\ldots,j_d=0,1} C^{(j)}_{I^{(n)}_j} + 1. \tag{1.104}$$

Here $(j_1, \ldots, j_d)$ is the binary representation of j, i.e.

$$j = \sum_{i=1}^{d} j_i 2^{d-i}. \tag{1.105}$$

In (1.104) $Y = (Y_1, \ldots, Y_s), U = (U_1, \ldots, U_d), (C_i^{(0)}), \ldots, (C_i^{2^d-1})$ are independent, Y and U are uniformly distributed on $[0,1]^s$ and $[0,1]^d$ respectively, $C_i^{(k)} \stackrel{\mathcal{D}}{=} C_i$ and $I^{(n)}$ is conditionally given $U = u$ multinomial $M(n-1, \langle u \rangle)$ distributed.

For d-dimensional quadtrees a first order asymptotic is known only for the mean of C_n. In Flajolet, Gonnet, Puech, and Robson (1993) the asymptotic expansion

$$\mathbb{E}\, C_n \sim \gamma_{s,d} n^{\alpha-1} \tag{1.106}$$

is proved. Here $\gamma_{s,d}$ is a positive constant which in principle can be approximated numerically and $\alpha \in (1,2)$ is the unique solution of the indicial equation

$$\alpha^{d-s}(\alpha+1)^s = 2^d. \tag{1.107}$$

An asymptotic for the variance of C_n has not been known so far. We will prove later that

$$\mathrm{Var}(C_n) \sim \beta_{s,d} n^{2\alpha-2}, \tag{1.108}$$

where $\beta_{s,d}$ has an explicit representation in terms of α and $\gamma_{s,d}$.
The normalized number of traversed nodes

$$X_n := \frac{C_n - \mathbb{E}\, C_n}{n^{\alpha-1}} \tag{1.109}$$

with α given by (1.107) satisfies the modified recursion

$$\begin{aligned} X_n \stackrel{\mathcal{D}}{=} & \sum_{j_1,\dots,j_s=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) \\ & \times \sum_{j_{s+1},\dots,j_d=0,1} \left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} (X^{(j)}_{I_j^{(n)}} + \gamma_{s,d}) - \gamma_{s,d} + o(1). \end{aligned} \tag{1.110}$$

We define

$$U_{j_1,\dots,j_d} := \prod_{\substack{1\le i\le d\\ j_i=0}} U_i \prod_{\substack{1\le i\le d\\ j_i=1}} (1-U_i) \tag{1.111}$$

for $j_1,\dots,j_d \in \{0,1\}$. By the convergence of the coefficients in (1.102) it seems reasonable that a distributional limit X of (X_n) is a solution of the limiting equation

$$X \stackrel{\mathcal{D}}{=} \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) U^{\alpha-1}_{j_1,\dots,j_d} (X^{(j)} + \gamma_{s,d}) - \gamma_{s,d}. \tag{1.112}$$

Therefore, we define the limiting operator

$$T : M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1)$$

$$T(\mu) :\stackrel{\mathcal{D}}{=} \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(Y,U) U^{\alpha-1}_{j_1,\dots,j_d} (Z^{(j)} + \gamma_{s,d}) - \gamma_{s,d}, \tag{1.113}$$

where $Y, U, Z^{(0)}, \dots, Z^{(2^d-1)}$ are independent, Y and U are uniformly distributed on $[0,1]^s$ and $[0,1]^d$ respectively, $Z^{(j)} \stackrel{\mathcal{D}}{=} \mu$ for $j = 0, \dots, 2^d - 1$.

Lemma 1.4.1 *The limiting operator $T : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. ℓ_2:*

$$\ell_2(T(\mu), T(\nu)) \le \xi\, \ell_2(\mu,\nu) \quad \textit{for all} \quad \mu,\nu \in M_{0,2}, \tag{1.114}$$

$$\xi = \frac{1}{\sqrt{\alpha^s(\alpha-1/2)^{d-s}}} < 1. \tag{1.115}$$

Proof: Obviously $\mathrm{Var}(T(\mu)) < \infty$. Since the summands in (1.113) are identically distributed we derive

$$\begin{aligned}
\mathbb{E}\, T(\mu) &= 2^d\, \mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i<U_i\}} \prod_{i=1}^{d} U_i^{\alpha-1}\right] \gamma_{s,d} - \gamma_{s,d} \\
&= 2^d\, \mathbb{E}\left[\mathbf{1}_{\{Y_1<U_1\}} U_1^{\alpha-1}\right]^s \mathbb{E}\left[U_1^{\alpha-1}\right]^{d-s} \gamma_{s,d} - \gamma_{s,d} \\
&= 2^d (\alpha+1)^{-s} \alpha^{-(d-s)} \gamma_{s,d} - \gamma_{s,d} \\
&= 2^d 2^{-d} \gamma_{s,d} - \gamma_{s,d} = 0, \qquad (1.116)
\end{aligned}$$

where the indicial equation (1.107) has been used. So $T : M_{0,2} \to M_{0,2}$ is well defined.

To prove contractivity let $\mu, \nu \in M_{0,2}$ and let $(W^{(k)}, Z^{(k)}), Y$ be independent, Y, U uniformly distributed on $[0,1]^s$ and $[0,1]^d$ respectively. Let $(W^{(k)}, Z^{(k)})$ be optimal ℓ_2–couplings of (μ, ν), i.e. $W^{(k)} \sim \mu$, $Z^{(k)} \sim \nu$, and $\ell_2^2(\mu,\nu) = \mathbb{E}\,(W^{(k)} - Z^{(k)})^2$ for $k = 0, \ldots, 2^d - 1$. Then using the independence properties and $\mathbb{E}\, W^{(k)} = \mathbb{E}\, Z^{(k)} = 0$ we conclude

$$\begin{aligned}
&\ell_2^2(T(\mu), T(\nu)) \\
&\leq \mathbb{E}\left[\sum_{j_1,\ldots,j_s=0,1} \mathbf{1}_{j_1,\ldots,j_s}(Y,U) \sum_{j_{s+1},\ldots,j_d=0,1} U_{j_1,\ldots,j_d}^{2\alpha-2} (Z^{(j)} - W^{(j)})^2\right] \\
&= 2^d\, \mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i<U_i\}} U_i^{2\alpha-2} \prod_{i=s+1}^{d} U_i^{2\alpha-2}\right] \ell_2^2(\mu,\nu) \\
&= 2^d (2\alpha)^{-s} (2\alpha-1)^{-(d-s)} \ell_2^2(\mu,\nu) \\
&= \frac{1}{\alpha^s (\alpha - 1/2)^{d-s}} \ell_2^2(\mu,\nu). \qquad (1.117)
\end{aligned}$$

This implies assertion (1.114). Since $\alpha \in (1,2)$ we have $2\alpha > \alpha+1$ and $2\alpha - 1 > \alpha$ which together with (1.107) yields

$$\begin{aligned}
\xi &= \left(\alpha^s (\alpha - 1/2)^{d-s}\right)^{-1/2} = \left(\frac{2^d}{(2\alpha)^s (2\alpha-1)^{d-s}}\right)^{1/2} \\
&< \frac{2^d}{(\alpha+1)^s \alpha^{d-s}} = 1. \qquad (1.118)
\end{aligned}$$

■

The representation of the limiting operator T can be simplified. Denote

$$\overline{U}_{j_{s+1},\dots,j_d} := \prod_{\substack{s+1\le i\le d\\ j_i=0}} U_i \prod_{\substack{s+1\le i\le d\\ j_i=1}} (1-U_i). \tag{1.119}$$

for $j_{s+1},\dots,j_d \in \{0,1\}$. Then

$$T(\mu) \stackrel{\mathcal{D}}{=} \prod_{i=1}^{s} U_i^{\frac{\alpha-1}{2}} \left(\sum_{j_{s+1},\dots,j_d=0,1} \overline{U}^{\alpha-1}_{j_{s+1},\dots,j_d} (X^{(j_{s+1},\dots,j_d)} + \gamma_{s,d}) \right) - \gamma_{s,d}, \tag{1.120}$$

where $\{U, X^{(j_{s+1},\dots,j_d)} : j_{s+1},\dots,j_d = 0,1\}$ is independent, U uniformly distributed on $[0,1]^d$ and $X^{(j_{s+1},\dots,j_d)} \stackrel{\mathcal{D}}{=} \mu$ for all $j_{s+1},\dots,j_d = 0,1$. The proof follows from an elementary calculation observing that some sets of the indicator functions in (1.112) resp. (1.103) are disjoint and $\sqrt{U}$ has the density $2x$ for $0 \le x \le 1$. With an additional translation this can be written more concisely. X is a fixed point of T in $M_{0,2}$ if and only if $\widetilde{X} := X + \gamma_{s,d}$ is a fixed point of the operator $\widetilde{T}$ on $M_{\gamma_{s,d},2}$ given by

$$\widetilde{T}(\mu) :\stackrel{\mathcal{D}}{=} \prod_{i=1}^{s} U_i^{\frac{\alpha-1}{2}} \sum_{j_{s+1},\dots,j_d=0,1} \overline{U}^{\alpha-1}_{j_{s+1},\dots,j_d} \widetilde{X}^{(j_{s+1},\dots,j_d)}. \tag{1.121}$$

In (1.121) again $\{U, \widetilde{X}^{(j_{s+1},\dots,j_d)} : j_{s+1},\dots,j_d = 0,1\}$ is independent, U uniformly distributed on $[0,1]^d$ and $\widetilde{X}^{(j_{s+1},\dots,j_d)} \stackrel{\mathcal{D}}{=} \mu \in M_{\gamma_{s,d},2}$.

Theorem 1.4.2 (Limit Theorem for Partial Match Query in Quadtrees) *The normalized number X_n of nodes traversed during a partial match query in a random d-dimensional quadtree with $1 \le s \le d-1$ components specified converges w.r.t. ℓ_2 to the unique fixed point X in $M_{0,2}$ of the limiting operator T given in (1.113), i.e.*

$$\ell_2(X_n, X) \to 0. \tag{1.122}$$

The translated limiting distribution $\widetilde{X} := X + \gamma_{s,d}$ is the unique fixed point in $M_{\gamma_{s,d},2}$ of the operator

$$\widetilde{T}(\mu) \stackrel{\mathcal{D}}{=} \prod_{i=1}^{s} U_i^{\frac{\alpha-1}{2}} \sum_{j_{s+1},\dots,j_d=0,1} \overline{U}^{\alpha-1}_{j_{s+1},\dots,j_d} \widetilde{X}^{(j_{s+1},\dots,j_d)} \tag{1.123}$$

given in (1.121).

Proof: Using (1.102), (1.110), and (1.114) the proof of Theorem 1.1.2 can be adapted to this situation. With $a_n := \ell_2^2(X_n, X)$ analogously to (1.27)–(1.31) the recursion

$$a_n \le 2^d \sum_{j=0}^{n-1} \mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i<U_i\}} \mathbf{1}_{\{I_1^{(n)}=j\}} (j/n)^{2\alpha-2} + o(1)\right] a_i + o(1) \qquad (1.124)$$

can be derived. Then as in (1.33)

$$\begin{aligned} a_n &\le 2^d \left(\mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i<U_i\}} U_i^{2\alpha-2} \prod_{i=s+1}^{d} U_i^{2\alpha-2}\right] + o(1)\right) \sup_{1\le i\le n-1} a_i + o(1) \\ &= (\xi^2 + o(1)) \sup_{1\le i\le n-1} a_i + o(1) \end{aligned} \qquad (1.125)$$

with ξ given in (1.115). This implies that $(a_n)_{n\in\mathbb{N}}$ is bounded. The convergence then follows as in (1.34). ■

Convergence in the ℓ_2-metric implies convergence of the second moments. Therefore a first order asymptotic of the variance of the number of traversed nodes C_n can be derived in dimension d with $1 \le s \le d-1$ components in the search pattern specified. We use the simplified form of the fixed point equation (1.121).

Corollary 1.4.3 (Variance of Partial Match Query in Quadtrees) *The variance of the limiting distribution for the normalized number of traversed nodes during a partial match query in a random d-dimensional quadtree with $1 \le s \le d-1$ components specified is given by*

$$\beta_{s,d} := \left[\frac{((2\alpha-1)B(\alpha,\alpha)+1)^{d-s}-1}{\alpha^s(\alpha-1/2)^{d-s}-1} - 1\right]\gamma_{s,d}. \qquad (1.126)$$

The variance of the number of traversed nodes C_n satisfies

$$Var(C_n) \sim \beta_{s,d} n^{2\alpha-2}. \qquad (1.127)$$

The constants α and $\gamma_{s,d}$ are given by (1.106)–(1.107) and $B(\,\cdot\,,\,\cdot\,)$ denotes the Eulerian beta integral.

Proof: Observe that

$$\mathrm{Var}(X) = \mathrm{Var}(\widetilde{X}) = \mathbb{E}\,\widetilde{X}^2 - \gamma_{s,d}^2 \qquad (1.128)$$

where $\widetilde{X}$ is the fixed point of the operator in (1.121). From (1.121) we obtain

$$\mathbb{E}\,\widetilde{X}^2 = \mathbb{E}\left[\prod_{i=1}^{s} U_i^{\alpha-1} \sum_{\substack{j_{s+1},\dots,j_d=0,1\\ k_{s+1},\dots,k_d=0,1}} \overline{U}_{j_{s+1},\dots,j_d}^{\alpha-1} \overline{U}_{k_{s+1},\dots,k_d}^{\alpha-1} \widetilde{X}^{(j_{s+1},\dots,j_d)} \widetilde{X}^{(k_{s+1},\dots,k_d)}\right]$$

$$= \alpha^{-s}\Bigg[\sum_{\forall\, i:\, j_i=k_i} \mathbb{E}\,\overline{U}^{2\alpha-2}_{j_{s+1},\dots,j_d}\,\mathbb{E}\,\widetilde{X}^2 + \sum_{\exists\, i:\, j_i\neq k_i} \mathbb{E}\left[\overline{U}^{\alpha-1}_{j_{s+1},\dots,j_d}\overline{U}^{\alpha-1}_{k_{s+1},\dots,k_d}\right](\mathbb{E}\,\widetilde{X})^2\Bigg]. \tag{1.129}$$

The expectations of the occurring $\overline{U}$s can be calculated explicitly:

$$\mathbb{E}\,\overline{U}^{2\alpha-2}_{j_{s+1},\dots,j_d} = (2\alpha-1)^{-(d-s)} \tag{1.130}$$

and for $(j_{s+1},\dots,j_d),(k_{s+1},\dots,k_d)$ and

$$h := \operatorname{card}\{s+1\le i\le d : j_i\neq k_i\},$$

$$\begin{aligned}\mathbb{E}\left[\overline{U}^{\alpha-1}_{j_{s+1},\dots,j_d}\overline{U}^{\alpha-1}_{k_{s+1},\dots,k_d}\right] &= \left(\mathbb{E}\,U_1^{2\alpha-2}\right)^{d-s-h}\left(\mathbb{E}\,(U_1(1-U_1))^{\alpha-1}\right)^h \\ &= (2\alpha-1)^{-(d-s-h)}(B(\alpha,\alpha))^h. \end{aligned}\tag{1.131}$$

With (1.131) in (1.129) we derive

$$\mathbb{E}\,\widetilde{X}^2 = \alpha^{-s}\Bigg[2^{d-s}(2\alpha-1)^{-(d-s)}\,\mathbb{E}\,\widetilde{X}^2 + \sum_{h=1}^{d-s} 2^{d-s}\binom{d-s}{h}(2\alpha-1)^{-(d-s-h)}(B(\alpha,\alpha))^h\gamma^2_{s,d}\Bigg]. \tag{1.132}$$

Using the binomial formula it follows

$$\begin{aligned}&\mathbb{E}\,\widetilde{X}^2\left(1-\alpha^{-s}\left(\frac{2}{2\alpha-1}\right)^{d-s}\right)\\ &= \alpha^{-s}2^{d-s}\left[\left(B(\alpha,\alpha)+\frac{1}{2\alpha-1}\right)^{d-s}-\left(\frac{1}{2\alpha-1}\right)^{d-s}\right]\gamma^2_{s,d}.\end{aligned}\tag{1.133}$$

A simplification leads to

$$\mathbb{E}\,\widetilde{X}^2 = \frac{((2\alpha-1)B(\alpha,\alpha)+1)^{d-s}-1}{\alpha^s(\alpha-1/2)^{d-s}-1}\gamma^2_{s,d}. \tag{1.134}$$

Together with (1.128) this implies the first assertion.
By convergence of the second moments of X_n we finally conclude

$$\begin{aligned}\operatorname{Var}(C_n) &= \operatorname{Var}(n^{\alpha-1}X_n) = \operatorname{Var}(X_n)n^{2\alpha-2} = (\operatorname{Var}(X)+o(1))n^{2\alpha-2}\\ &\sim \beta_{s,d}n^{2\alpha-2}.\end{aligned}\tag{1.135}$$

■

The standard quadtree in dimension $\mathbf{d} = 2$:

For the 2-dimensional quadtree the constants $\gamma_{1,2}$ and α in (1.106) are explicitly known: Flajolet, Gonnet, Puech, and Robson (1993) derive

$$\mathbb{E}\, C_n \sim \gamma n^{\alpha-1} \tag{1.136}$$

with

$$\alpha = \frac{\sqrt{17}-1}{3} \quad \text{and} \quad \gamma = \frac{\Gamma(2\alpha)}{2\Gamma^3(\alpha)}. \tag{1.137}$$

Therefore we can state our results in dimension 2 more explicitly. The contraction factor ξ in (1.115) is given by

$$\xi = \frac{2}{\sqrt{19-3\sqrt{17}}} = 0.776\ldots \tag{1.138}$$

The simplified limiting operator reads

$$T(\mu) \stackrel{\mathcal{D}}{=} U^{\frac{\alpha-1}{2}}\{V^{\alpha-1}(Z^{(1)}+\gamma) + (1-V)^{\alpha-1}(Z^{(2)}+\gamma)\} - \gamma \tag{1.139}$$

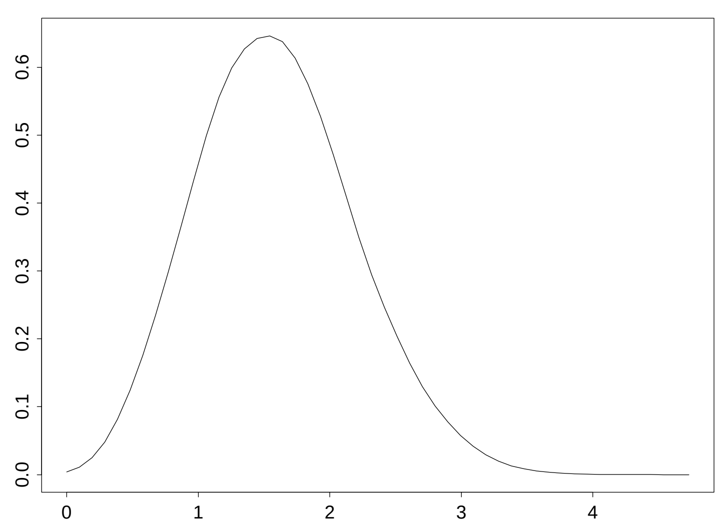
with $U, V, Z^{(1)}, Z^{(2)}$ being independent, U, V uniformly distributed on $[0,1]$, and $Z^{(1)}, Z^{(2)} \stackrel{\mathcal{D}}{=} \mu$. By the additional translation it follows that X is the weak limit of the normalized cost X_n of a partial match query in a random 2-dimensional quadtree if and only if $\widetilde{X} := X + \gamma$ is the fixed point of

$$\widetilde{T}(\mu) :\stackrel{\mathcal{D}}{=} U^{\frac{\alpha-1}{2}}(V^{\alpha-1}\widetilde{Z}^{(1)} + (1-V)^{\alpha-1}\widetilde{Z}^{(2)}) \tag{1.140}$$

in $M_{\gamma,2}$.

Since the constants α, γ have explicit representations we can simulate an approximation of the limiting distribution iterating the limiting operator $\widetilde{T}$.

estimated density of the translated limiting distribution



The plot was produced by iterating the translated limiting operator $\widetilde{T}$ 10 times starting with δ_γ, the Dirac measure in γ. We produced 15 000 samples of $\widetilde{T}^{10}(\delta_\gamma)$ and applied a standard smoothing-routine of S-Plus on the histogram of the data.

1.5 Moments, tail, and large deviation

The techniques developed by Rösler (1991, 1992) to obtain results on the existence and convergence of Laplace transforms for the scaled running time of the quicksort algorithm can be applied to the recursions of the partial match query type studied in the previous sections.

Theorem 1.5.1 (Laplace transforms) *The limit X of the normalized number of nodes traversed during a partial match query in a random K-d tree, random K-d-t tree or random relaxed K-d tree with $1 \leq s \leq K-1$ components specified or in a random quadtree with $1 \leq s \leq d-1$ components specified has a finite Laplace transform in some neighborhood of 0,*

$$\mathbb{E}\exp(\lambda X) < \infty \quad \text{for all} \quad \lambda \in (-\lambda_0, \lambda_0). \tag{1.141}$$

Assume

$$0 < \frac{s}{K} \leq \frac{\ln(4/3)}{\ln(5/3)} = 0.563\ldots \quad \text{for the } K\text{-d tree}, \tag{1.142}$$

$$0 < \frac{s}{K} \leq \frac{\ln\left(\frac{4+2t}{3+2t}\right)^{\overline{t+1}}}{\ln\left(\frac{5+2t}{3+2t}\right)^{\overline{t+1}}} \quad \text{for the } K\text{-d-t tree}, \tag{1.143}$$

$$0 < \frac{s}{K} \leq 0.625 \quad \text{for the random relaxed } K\text{-d tree}, \tag{1.144}$$

$$0 < \frac{s}{d} \leq \frac{\ln(4/3)}{\ln(5/3)} = 0.563\ldots \quad \text{for the quadtree}. \tag{1.145}$$

Then existence and convergence of the Laplace transform hold on the whole real line:

$$\mathbb{E}\exp(\lambda X_n) \to \mathbb{E}\exp(\lambda X) \quad \text{for all} \quad \lambda \in \mathbb{R}. \tag{1.146}$$

Proof: The proof is given here for the quadtree. The other cases can be deduced analogously. Observe that the recursions for X_n and X given in (1.110) and (1.112) can be written in the form

$$X_n \stackrel{\mathcal{D}}{=} \sum_{j_1,\ldots,j_d=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y)\left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} X^{(j)}_{I_j^{(n)}} + C_n(U,Y,I^{(n)}) \tag{1.147}$$

and

$$X \stackrel{\mathcal{D}}{=} \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) U_{j_1,\dots,j_d}^{\alpha-1} X^{(j)} + C(U,Y) \tag{1.148}$$

with

$$C_n(U,Y,I^{(n)}) = \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) \left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} \gamma_{s,d} - \gamma_{s,d} + o(1) \tag{1.149}$$

and

$$C(U,Y) = \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) U_{j_1,\dots,j_d}^{\alpha-1} \gamma_{s,d} - \gamma_{s,d}. \tag{1.150}$$

The distributions and (in-)dependencies are as in (1.110) and (1.112). The recursion (1.148) satsfies the conditions of Theorem 6 in Rösler (1992) with

$$T_j = \mathbf{1}_{j_1,\dots,j_s}(U,Y) \left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1}. \tag{1.151}$$

This implies the existence of a neighborhood $(-\lambda_0, \lambda_0)$ where X has finite Laplace transform.

For the second assertion note that

$$\mathbb{E}\, C_n(U,Y,I^{(n)}) = 0 \quad \text{for all} \quad n \in \mathbb{N} \tag{1.152}$$

since the X_n and $X_{I_j^{(n)}}^{(j)}$ in (1.147) are centered. Define

$$V_n := \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) \left(\frac{I_j^{(n)}}{n}\right)^{2\alpha-2} - 1. \tag{1.153}$$

It is $\sum_{j=0}^{2^d-1} I_j^{(n)} = n-1$. The condition (1.145) and the indicial equation (1.107) imply $\alpha \geq 3/2$. Thus

$$V_n < 0 \quad \text{for all} \quad n \in \mathbb{N}. \tag{1.154}$$

The convergence of the coefficients in (1.102) implies

$$\mathbb{E}\, V_n \longrightarrow \mathbb{E} \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) U_{j_1,\dots,j_d}^{2\alpha-2} - 1 = \xi - 1 < 0 \tag{1.155}$$

with ξ given in (1.115). This yields

$$\sup_{n\in\mathbb{N}} \mathbb{E}\, V_n < 0. \tag{1.156}$$

From the representation (1.149) of $C_n(U, Y, I^{(n)})$ it is obvious that

$$\sup_{n\in\mathbb{N}} \|C_n\|_\infty < \infty. \tag{1.157}$$

The properties (1.152), (1.154), (1.156) and (1.157) are sufficient to obtain

$$\mathbb{E}\exp(\lambda X_n) \longrightarrow \mathbb{E}\exp(\lambda X), \quad \lambda \in \mathbb{R}. \tag{1.158}$$

as in Lemma 4.1 and Theorem 4.2 in Rösler (1991). ■

In particular Theorem 1.5.1 implies exponential tails and the existence of all moments of the limiting distributions. Under condition (1.145) (resp. (1.142), (1.143), (1.144)) additionally convergence of all moments follows and an estimate for large deviations of the (unscaled) cost C_n can be established: For all $\lambda > 0$ there exists a $c_\lambda > 0$ so that for any sequence (a_n) of positive, real numbers holds:

$$\mathbb{P}(C_n \geq a_n) \leq c_\lambda \exp\left(-\lambda \frac{a_n}{n^{\alpha-1}}\right). \tag{1.159}$$

The existence of densities of the limiting distributions with respect to the Lebesgue measure can be deduced following the scheme of Theorem 2.1 in Tan and Hadjicostas (1995) for the limiting distribution of the running time of the Quicksort algorithm. The translated limit distributions (given by the operators (1.38), (1.95), (1.121) and analoguosly for the K-d-t tree) are supported by $[0, \infty)$. The densities are positive almost everywhere on $[0, \infty)$ (cf. Theorem 2.4 in Tan and Hadjicostas (1995)).

Chapter 2

Internal path length

Trees are fundamental data structures for the purpose of sorting and searching in a file. There are several characteristics for measuring the performance of a certain tree. The *depth* of a node in a tree is the number of nodes from the root down to this node. It indicates the effort to insert this node into the tree. The *height* of a tree is the maximal depth of its nodes, which measures the worst case performance for inserting a node into the tree. The *internal path length* is the sum of all depths of the internal nodes in the tree. The internal path length, therefore, is a indicator for the total cost for building up the tree from the data.

In this chapter the internal path length of several trees, which are based on key comparisons is under consideration. The data are taken independently and identically distributed from some distribution to build up the tree. We are interested in limit laws for the normalized internal path length.

Some common trees are the random binary search tree, the random m-ary search tree, the random quadtree and the random median-of-$(2k+1)$ tree. In order to derive a uniform limit theorem for the internal path length of a more general type of tree which includes in particular these trees we consider the *random split tree model* introduced in Devroye (1998). The random split tree is a general model of a tree which contains all the trees mentioned above as well as many other variants of trees. A related model of a general class of random trees is discussed in Aldous (1996).

In the first section of this chapter the notion of the random split tree is recalled. Then the limit law for a large class of split trees is given. In the final section this limit theorem is specialized to the random quadtree and the random m-ary search tree, for which the limit law of the internal path length has been unknown so far.

2.1 The random split tree

We briefly recall how the split tree works: Given are a fixed branch factor b, the vertex capacity s and for the definition of the distribution process two additional integers s_0 and s_1 with

$$0 < s, \;\; 0 \le s_0 \le s, \;\; 0 \le bs_1 \le s + 1 - s_0. \tag{2.1}$$

Furthermore a random splitting vector $\mathcal{V} = (V_1, \dots, V_b)$ of (random) probabilities, $\sum V_k = 1, V_k \ge 0$, is given. Now, the corresponding random split tree is constructed in the following way. At each node of an empty *skeleton tree* with branch factor b (this is an infinite rooted tree with b subtrees at each node) an independent copy of $\mathcal{V}$ is attached. Each node holds at most s items. Initially, there are no items in the tree. Items are added to the tree rooted at u as follows. Let $(V_1, \dots, V_b)$ be the splitting vector at node u. If u is not a leaf choose subtree i at random according to the probabilities $(V_1, \dots, V_b)$ and add this item to the ith subtree. If u is a leaf with less than s (vertex capacity) items, only add this additional item to the leaf. If u is a leaf already holding s items we have to distribute the $s + 1$ items: Place s_0 randomly selected items at u, then send s_1 randomly selected items to each of the subtrees. The remaining $s + 1 - s_0 - bs_1$ items are sent down to the subtrees independently at random according to the splitting probabilities $(V_1, \dots, V_b)$. This process may have to be repeated several times if $s_0 = 0$.

The random binary search tree built up from independent, uniformly on $[0, 1]$ distributed data for example corresponds to the random split tree with vertex capacity $s = 1$, branch factor $b = 2$ and splitting vector $\mathcal{V} = (U, 1 - U)$, where U is uniformly distributed on $[0, 1]$. Further $s_0 = 1$ and $s_1 = 0$. The parameters for fitting the other common trees into the model of a random split tree are given in Devroye (1998).

Devroye gives a universal law of large numbers and a universal limit law for the depth D_n of the nth inserted item as well as a general law of large numbers for the height H_n of the split tree with n items. Without loss of generality it can be assumed that the splitting vector $\mathcal{V} = (V_1, \dots, V_b)$ has identically distributed components (see Devroye (1998)). Then a r.v. V with $V \sim V_1$ is called a *splitter*. The asymptotic behavior of the depth D_n only depends on

$$\mu := b\,\mathbb{E}\,[V \ln(1/V)] \quad \text{and} \tag{2.2}$$

$$\sigma^2 := b\,\mathbb{E}\,[V \ln^2 V] - \mu^2. \tag{2.3}$$

The asymptotic of the height is related additionally to the moments of the splitter. Observe that for the analysis of the depth and the law of large numbers of the height there is no knowledge of the joint distribution of the splitting vector

$\mathcal{V} = (V_1, \ldots, V_b)$ needed. Only the marginal distribution of the splitter is used for these asymptotics. We will see, that the limit law for the internal path length depends on the joint distribution of $(V_1, \ldots, V_b)$.

2.2 Internal path length in split trees

For the distributional analysis of the internal path length in a random split tree the method introduced by Rösler (1991) for the analysis of the running time of the quicksort algorithm is extended. As it is well known the running time of quicksort is distributed as the internal path length of the random binary search tree. It seems difficult to extend the martingal method, which was used in Régnier (1989) for the analysis of quicksort.

In the following we assume that there are no empty internal nodes in the split tree, i.e. $s_0 \geq 1$. Furthermore assume that the splitting vector $\mathcal{V} = (V_1, \ldots, V_b)$ has identically distributed components which are continuous with respect to the Lebesgue measure. Now the splitting vector at the root of the tree is denoted by $\mathcal{V} = (V_1, \ldots, V_b)$. Denote by $I^{(n)}$ the vector of the cardinalities of the subtrees of the root of the random split tree with n keys inserted. For $n \geq s + 1$ according to the distribution process of the items we have

$$\mathbb{P}^{I^{(n)} \,|\, \mathcal{V}=(v_1,\ldots,v_b)} = (\delta_{s_1} \otimes \ldots \otimes \delta_{s_1}) * M(n - s_0 - bs_1, v_1, \ldots, v_b). \tag{2.4}$$

This means that given $\mathcal{V}$ the cardinalities are distributed as the fixed vector $(s_1, \ldots, s_1)$ plus a multinomial r.v. with the indicated parameters. In particular this implies a weak law of large numbers for $I^{(n)}$:

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \mathcal{V} = (V_1, \ldots, V_b). \tag{2.5}$$

Let Y_n denote the internal path length of the random split tree with n items inserted, i.e. Y_n is the sum of the depths of the items in the tree. The depth of the root is defined to be one; $Y_0 := 0$. Since the subtrees of the random split tree are given their cardinalities again distributed as random split trees with the same parameters and independent of each other the following recursion for the internal path length holds in distribution

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{k=1}^{b} Y^{(k)}_{I^{(n)}_k} + n. \tag{2.6}$$

Here $(Y_i^{(k)})$ are independent copies of (Y_i) and $\{(Y_i^{(k)}) : k = 1, \ldots, b\}, I^{(n)}$ are independent. For the scaling we assume that the mean of Y_n admits an expansion of the form

$$\mathbb{E}\, Y_n = \mu^{-1} n \ln n + dn + o(n) \tag{2.7}$$

with $\mu > 0$ given in (2.2) and $d \in \mathbb{R}$. In the next section examples of trees of this type and a general approach to a derivation of such an expansion are discussed. It is proved later that for these split trees the variance of Y_n is asymptotically of the form $\sim vn^2$. Therefore the right normalization of Y_n is given by

$$X_n := \frac{Y_n - \mathbb{E}\, Y_n}{n}. \tag{2.8}$$

A straightforward calculation using (2.6) leads to the modified recursion

$$X_n \stackrel{\mathcal{D}}{=} \sum_{k=1}^{b} \frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} + C_n(I^{(n)}) \tag{2.9}$$

where $(X_i^{(k)})$ are independent copies of X_i, further $\{(X_i^{(k)}),\ k = 1, \ldots, b\}, I^{(n)}$ are independent and

$$C_n(i) := 1 + \frac{1}{n}\left(\sum_{k=1}^{b} \mathbb{E}\, Y_{i_k} - \mathbb{E}\, Y_n\right) \tag{2.10}$$

for $i = (i_1, \ldots, i_b)$ with $\sum i_k = n - s_0$. In order to obtain a limiting form of the recursion (2.9) we introduce the simplex

$$T_b := \left\{x \in [0,1]^b : \sum_{k=1}^{b} x_k = 1\right\} \tag{2.11}$$

and the entropy functional

$$C : T_b \to \mathbb{R}, \ \ C(x) := 1 + \mu^{-1} \sum_{k=1}^{b} x_k \ln x_k \tag{2.12}$$

where $x \ln x$ is defined to be 0 for $x = 0$. Let

$$G_n := \left\{(i_1, \ldots, i_b) \in \mathbb{N}^b : \sum_{k=1}^{b} i_k = n - s_0\right\} \tag{2.13}$$

denote the domain of C_n, then as in Rösler (1991) C approximates C_n in the following sense:

Lemma 2.2.1 *Let $(z^{(n)})$ be a sequence, $z^{(n)} \in G_n$ such that $z^{(n)}/n \to z \in (0,1]^b$, then*

$$\lim_{n\to\infty} C_n(z^{(n)}) = C(z). \tag{2.14}$$

Furthermore

$$\sup_{n\in\mathbb{N}} \|C_n\|_\infty < \infty. \tag{2.15}$$

Proof: Let $(z^{(n)})$ be a sequence, $(z^{(n)}) \in G_n$ such that $z^{(n)}/n \to z \in (0,1]^b$. Using the expansion (2.7) of the expectation of Y_n,

$$\mathbb{E}\, Y_n = \mu^{-1} n \ln n + dn + R_n \quad \text{with} \quad R_n/n = o(1),$$

we obtain

$$\begin{aligned} C_n(z^{(n)}) &= 1 + \frac{1}{n}\left(\sum_{k=1}^{b} \mathbb{E}\, Y_{z_k^{(n)}} - \mathbb{E}\, Y_n\right) && (2.16)\\ &= 1 + \frac{1}{n}\left(\sum_{k=1}^{b}\left(\mu^{-1} z_k^{(n)} \ln z_k^{(n)} + d z_k^{(n)} + R_{z_k^{(n)}}\right) - \mu^{-1} n \ln n - dn - R_n\right). \end{aligned}$$

Since $\sum_{k=1}^{b} z_k^{(n)} = n - s_0$ by definition of G_n,

$$\sum_{k=1}^{b} \mu^{-1} z_k^{(n)} \ln z_k^{(n)} - \mu^{-1} n \ln n = \sum_{k=1}^{b} \mu^{-1} z_k^{(n)} \ln \frac{z_k^{(n)}}{n} - \mu^{-1} s_0 \ln n. \qquad (2.17)$$

With (2.16) and (2.17) we derive

$$\begin{aligned} C_n(z^{(n)}) &= 1 + \frac{1}{n}\Bigg(\sum_{k=1}^{b}\left(\mu^{-1} z_k^{(n)} \ln \frac{z_k^{(n)}}{n} + R_{z_k^{(n)}}\right) && (2.18)\\ &\qquad - \mu^{-1} s_0 \ln n - s_0 d - R_n\Bigg) \\ &= C(z^{(n)}/n) + \frac{1}{n}\sum_{k=1}^{b} R_{z_k^{(n)}} - \frac{1}{n}\left(\mu^{-1} s_0 \ln n - s_0 d - R_n\right). \end{aligned}$$

Now observe that

$$2\alpha := \min_{1 \le k \le b} z_k > 0 \qquad (2.19)$$

since $z \in (0,1]^b$. This implies $z_k^{(n)} \ge \alpha n$ for $1 \le k \le b$ and n sufficiently large. Let $\overline{R}_n := \sup_{i \ge n} |R_i|$. Then $(\overline{R}_n)_{n\in\mathbb{N}}$ is decreasing and $\overline{R}_n/n \to 0$. For n sufficiently large it follows (denoting by $\lfloor x \rfloor$ the integer part of x for $x \ge 0$)

$$\begin{aligned} \left|\frac{1}{n}\sum_{k=1}^{b} R_{z_k^{(n)}}\right| &\le \frac{1}{n}\sum_{k=1}^{b} \overline{R}_{z_k^{(n)}} \\ &\le \frac{1}{n}\sum_{k=1}^{b} \overline{R}_{\lfloor \alpha n \rfloor} \\ &\le b\alpha \frac{1}{\lfloor \alpha n \rfloor} \overline{R}_{\lfloor \alpha n \rfloor} \to 0 \quad \text{for} \quad n \to \infty. && (2.20) \end{aligned}$$

The third summand of (2.18) also tends to zero. So (2.18) implies

$$C_n(z^{(n)}) = C(z^{(n)}/n) + o(1).$$

Therefore, continuity of C and the triangle inequality imply

$$\left|C_n(z^{(n)}) - C(z)\right| \leq \left|C_n(z^{(n)}) - C(z^{(n)}/n)\right| + \left|C(z^{(n)}/n) - C(z)\right| \longrightarrow 0.$$

In order to get an estimate for $C_n(z^{(n)})$ uniformly in $z^{(n)} \in G_n$ let

$$L := \sup_{n\in\mathbb{N}} |R_n/n| < \infty.$$

Then (2.18) implies

$$\begin{aligned} |C_n(z^{(n)})| &\leq |C(z^{(n)}/n)| + bL + o(1) \\ &\leq \|C\|_\infty + bL + o(1). \end{aligned} \tag{2.21}$$

The second claim follows. ■

Lemma 2.2.1 and the convergence of the coefficients in (2.5) suggest that a limit X of (X_n) is a solution of the limiting equation

$$X \stackrel{\mathcal{D}}{=} \sum_{k=1}^{b} V_k X^{(k)} + C(\mathcal{V}) \tag{2.22}$$

where $X^{(k)}$ are i.i.d. copies of X further $\{X^{(k)}, k = 1, \ldots b\}, \mathcal{V}$ are independent and $\mathcal{V}$ is a splitting vector. Define the random affine operator

$$T : M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1), \quad T(\mu) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{b} V_k Z^{(k)} + C(\mathcal{V}) \tag{2.23}$$

where $(Z^{(k)}), \mathcal{V}$ are independent, $Z^{(k)} \sim \mu$ and $\mathcal{V}$ is a splitting vector.

Lemma 2.2.2 *$T : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. ℓ_2:*

$$\ell_2(T(\mu), T(\nu)) \leq \|\mathcal{V}\|_2 \, \ell_2(\mu, \nu) \qquad \text{for all } \mu, \nu \in M_{0,2}. \tag{2.24}$$

with

$$\|\mathcal{V}\|_2 = \left(\sum_{k=1}^{b} V_k^2\right)^{1/2} < 1. \tag{2.25}$$

Proof: Obviously Var $(T(\mu)) < \infty$ and using (2.2)

$$\mathbb{E}\, T(\mu) = \mathbb{E}\, C(\mathcal{V}) = 1 + \mu^{-1} b\, \mathbb{E}\,[V_1 \ln V_1] = 0. \tag{2.26}$$

so T is a well defined mapping $T : M_{0,2} \to M_{0,2}$.

To prove contractivity let $\mu, \nu \in M_{0,2}$ and let $(W^{(k)}, Z^{(k)}), \mathcal{V}$ be independent, $\mathcal{V}$ a splitting vector. Let $(W^{(k)}, Z^{(k)})$ be optimal ℓ_2–couplings of (μ, ν), i.e. $W^{(k)} \sim \mu$, $Z^{(k)} \sim \nu$ and $\ell_2^2(\mu,\nu) = \mathbb{E}\,(W^{(k)} - Z^{(k)})^2$. Then using the independence properties and $\mathbb{E}\, W^{(k)} = \mathbb{E}\, Z^{(k)} = 0$

$$\begin{aligned}
\ell_2^2(T(\mu), T(\nu)) &= \ell_2^2\left(\sum_{k=1}^b V_k W^{(k)} + C(V), \sum_{k=1}^b V_k Z^{(k)} + C(V)\right) \\
&\le \mathbb{E}\left(\sum_{k=1}^b V_k (W^{(k)} - Z^{(k)})\right)^2 \\
&= \sum_{k=1}^b \mathbb{E}\left[V_k^2 (W^{(k)} - Z^{(k)})^2\right] \\
&= b\ \mathbb{E}\, V_1^2\ \ell_2^2(\mu,\nu) \\
&= \|\mathcal{V}\|_2^2\ \ell_2^2(\mu,\nu).
\end{aligned} \tag{2.27}$$

It is $\sum_{k=1}^b V_k = 1$ and $\mathbb{P}(V_k = 1) = 0$ since we assumed the existence of a density of V_k. This implies $\|\mathcal{V}\|_2^2 = \mathbb{E}\ \sum_{k=1}^b V_k^2 < \mathbb{E}\ \sum_{k=1}^b V_k = 1$. ■

By Banach's fixed point theorem T has a unique fixed point X in $M_{0,2}$.

Theorem 2.2.3 (Limit theorem for the internal path length in random split trees) *Let Y_n denote the internal path length of a random split tree with branch factor $b \geq 1$, vertex capacity $s \geq 1$, splitting vector $\mathcal{V} = (V_1, \ldots, V_b)$ and distribution parameters $s_0 \geq 1$ and $s_1 \geq 0$. Let V_i be identically distributed and continuous with respect to the Lebesgue measure. Assume*

$$\mathbb{E}\, Y_n = \mu^{-1} n \ln n + dn + o(n) \tag{2.28}$$

with $\mu = b\, \mathbb{E}\,[V_1 \ln(1/V_1)]$ and define

$$X_n := \frac{Y_n - \mathbb{E}\, Y_n}{n}. \tag{2.29}$$

Then the following holds:

(a) $\ell_2(X_n, X) \to 0,$ (2.30)

where X is the fixed point of T given in (2.23) in $M_{0,2}$,

(b) $Var(Y_n) \sim Var(X)\, n^2,$ (2.31)

(c) *exponential moments exist and converge:*

$$\mathbb{E}\exp(\lambda X_n) \to \mathbb{E}\exp(\lambda X) \quad \lambda \in \mathbb{R}, \tag{2.32}$$

(d) $\mathbb{P}(|Y_n - \mathbb{E}\,Y_n| \geq \varepsilon\,\mathbb{E}\,Y_n) = \mathcal{O}(n^{-k})$ *for all $k \in \mathbb{N}$.* (2.33)

Proof: `Ad (a)`: Let $X_n^{(k)} \sim X_n, X^{(k)} \sim X, 1 \leq k \leq b$ such that $(X_n^{(k)}, X^{(k)})$ are optimal couplings of X_n, X, i.e. $\ell_2^2(X_n, X) = \mathbb{E}\,(X_n^{(k)} - X^{(k)})^2$. Furthermore let $I^{(n)}$ be as in (2.4) and $\mathcal{V}$ a splitting vector. Finally assume that $((X_n^{(1)})_{n\in\mathbb{N}}, X^{(1)}), \ldots, ((X_n^{(b)})_{n\in\mathbb{N}}, X^{(b)}), (I^{(n)}/n, \mathcal{V})$ are independent. Then using the independence properties and $\mathbb{E}\,X^{(k)} = \mathbb{E}\,X_n^{(k)} = 0$ we obtain

$$\begin{aligned}
\ell_2^2(X_n, X) &= \ell_2^2\left(\sum_{k=1}^{b} \frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} + C_n(I^{(n)}), \sum_{k=1}^{b} V_k X^{(k)} + C(\mathcal{V})\right) \\
&\leq \mathbb{E}\left(\sum_{k=1}^{b}\left(\frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} - V_k X^{(k)}\right) + C_n(I^{(n)}) - C(\mathcal{V})\right)^2 \\
&= \mathbb{E}\left[\sum_{k=1}^{b}\left(\frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} - V_k X^{(k)}\right)^2 + \big(C_n(I^{(n)}) - C(\mathcal{V})\big)^2\right] \\
&= \sum_{k=1}^{b}\mathbb{E}\left(\frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} - V_k X^{(k)}\right)^2 + \mathbb{E}\,\big(C_n(I^{(n)}) - C(\mathcal{V})\big)^2.
\end{aligned} \tag{2.34}$$

By (2.5) respectively dominated convergence and Lemma 2.2.1 as $n \to \infty$

$$\mathbb{E}\,(I_k^{(n)}/n - V_k)^2 \to 0 \qquad \text{and} \tag{2.35}$$

$$\mathbb{E}\,(C_n(I^{(n)}) - C(\mathcal{V}))^2 \to 0. \tag{2.36}$$

For the first term of (2.34) consider

$$\begin{aligned}
&\mathbb{E}\left(\frac{I_k^{(n)}}{n} X^{(k)}_{I_k^{(n)}} - V_k X^{(k)}\right)^2 \\
&= \mathbb{E}\left(\frac{I_k^{(n)}}{n}\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right) + \left(\frac{I_k^{(n)}}{n} - V_k\right) X^{(k)}\right)^2
\end{aligned}$$

$$= \mathbb{E}\left(\frac{I_k^{(n)}}{n}\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right)\right)^2 + \mathbb{E}\left(\left(\frac{I_k^{(n)}}{n} - V_k\right)X^{(k)}\right)^2$$

$$+ 2\,\mathbb{E}\left[\frac{I_k^{(n)}}{n}\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right)\left(\frac{I_k^{(n)}}{n} - V_k\right)X^{(k)}\right]. \tag{2.37}$$

By independence and (2.35) the second term in (2.37) converges to zero. With the Cauchy-Schwarz-inequality and (2.35) the third term in its absolute value is estimated from above by

$$2\,\mathbb{E}\left[\left(\frac{I_k^{(n)}}{n}\right)^2\left(\frac{I_k^{(n)}}{n} - V_k\right)^2\left(X^{(k)}\right)^2\right]^{1/2}\left(\mathbb{E}\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right)^2\right)^{1/2}$$

$$\leq o(1)\,\mathbb{E}\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right)^2 + o(1). \tag{2.38}$$

The last inequality holds since both sides are o(1) if the expectation on the right hand side is less than 1. With (2.34)–(2.38) and denoting $a_n := \ell_2^2(X_n, X)$ we derive

$$a_n \leq b\,\mathbb{E}\left(\left(\frac{I_k^{(n)}}{n} + o(1)\right)\left(X^{(k)}_{I_k^{(n)}} - X^{(k)}\right)\right)^2 + b_n$$

$$= b\sum_{i=0}^{n-s_0}\mathbb{P}(\{(I_k^{(n)}/n) = (i/n)\})$$

$$\times\left((i/n)^2 + o(1)\right)\mathbb{E}\,(X_i^{(k)} - X^{(k)})^2 + b_n \tag{2.39}$$

where $b_n \to 0$ for $n \to \infty$. With (2.35) we conclude

$$a_n \leq b\sum_{i=0}^{n-s_0}\mathbb{P}(\{(I_k^{(n)}/n) = (i/n)\})\left((i/n)^2 + o(1)\right)\sup_{0\leq i\leq n-s_0} a_i + b_n$$

$$= (\|\mathcal{V}\|_2^2 + o(1))\sup_{0\leq i\leq n-s_0} a_i + b_n \tag{2.40}$$

which implies using $s_0 \geq 1$ that (a_n) is bounded. This implies as in Rösler (1991) that for a given $\varepsilon > 0$ there exists a n_0 such that for $n \geq n_0$

$$a_n \leq a + \varepsilon \qquad \text{with} \qquad a := \limsup_{n\to\infty} a_n \tag{2.41}$$

and the prefactor in (2.40) is uniformly less than a $\gamma < 1$. Therefore

$$a_n \leq b\sum_{i=0}^{n_0-1}\mathbb{P}\left(\frac{I_k^{(n)}}{n} = \frac{i}{n}\right)\left(\left(\frac{i}{n}\right)^2 + o(1)\right)a_i$$

$$
\begin{aligned}
& + b \sum_{i=n_0}^{n-s_0} \mathbb{P}\left(\frac{I_k^{(n)}}{n} = \frac{i}{n}\right) \left(\left(\frac{i}{n}\right)^2 + o(1)\right)(a+\varepsilon) + b_n \\
\leq \;& \gamma(a+\varepsilon) + o(1). \qquad (2.42)
\end{aligned}
$$

Then $0 \leq a = \limsup a_n \leq \gamma(a+\varepsilon)$ which implies $a = 0$.

`Ad (b):` The ℓ_2-convergence in (a) implies convergence of the second moments. This implies

$$
\begin{aligned}
\mathrm{Var}(Y_n) &= \mathrm{Var}(nX_n) = \mathrm{Var}(X_n)\, n^2 = (\mathrm{Var}(X) + o(1))\, n^2 \\
&\sim \mathrm{Var}(X)\, n^2. \qquad (2.43)
\end{aligned}
$$

`Ad (c):` Proceed as in Rösler (1991). Lemma 4.1 in Rösler (1991) for the special case of the random binary search tree can be extended to the general split tree:

$$
\forall\, L > 0\; \exists\, K_L > 0\, \forall\, n \in \mathbb{N}\; \forall\, \lambda \in [-L, L] : \; \mathbb{E}\, \exp(\lambda X_n) \leq \exp(\lambda^2 K_L). \qquad (2.44)
$$

In place of the random variable U_n in Röslers proof use

$$
W_n := \|I^{(n)}/n\|^2 - 1. \qquad (2.45)
$$

Then

(1) $-1 \leq W_n < 0$ for all $n \in \mathbb{N}$

(2) $\sup_{n\in\mathbb{N}} \mathbb{E}\, W_n < 0$

(3) $\sup_{n\in\mathbb{N}} \|C_n\|_\infty < \infty$ by Lemma 2.2.1.

For the proof of (2) note that $\mathbb{E}\, W_n < 0$ for all $n \in \mathbb{N}$ and $W_n \stackrel{\mathcal{D}}{\to} \|\mathcal{V}\|^2 - 1$ which implies by boundedness of W_n, $\mathbb{E}\, W_n \to \mathbb{E}\,(\|\mathcal{V}\|^2 - 1) < 0$. From (1)–(3) one obtains (2.44) as in Rösler (1991). The exponential bound in (2.44) implies uniform integrability of $\exp(\lambda X_n)$ which by part (a) yields (c).

`Ad (d):` Using the expansion (2.28) one deduces as in Corollary 4.3 in Rösler (1991): Let $k \in \mathbb{N}$ and $\varepsilon > 0$ given. With $\lambda := k\mu/\varepsilon$ in (c) we derive

$$
\begin{aligned}
\mathbb{P}(|Y_n - \mathbb{E}\, Y_n| \geq \varepsilon\, \mathbb{E}\, Y_n) &= \mathbb{P}(\exp(\lambda |X_n|) \geq \exp((\varepsilon\lambda\, \mathbb{E}\, X_n)/n)) \\
&\leq \frac{\mathbb{E}\, \exp(\lambda |X_n|)}{\exp(\lambda\varepsilon(\mu^{-1} \ln n + d + o(1)))} \\
&\leq (\mathbb{E}\, \exp(\lambda |X| - \lambda\varepsilon d) + o(1))\, n^{-\lambda\varepsilon\mu^{-1}} \\
&= \mathcal{O}(n^{-k}), \qquad (2.46)
\end{aligned}
$$

where Markov's inequality has been used. ■

2.3 Applications to special trees

The internal path length of some special cases of the random split tree have already been analyzed. For the random binary search tree the results of Theorem 2.2.3 have been given in the original paper of Rösler (1991). The random median–of–$(2k+1)$ search tree has been treated in Rösler (1999). This tree corresponds to the random split tree with branch factor $b = 2$, vertex capacity $s_0 = 1$, distribution parameters $s_1 = k, s = 2k$ and splitting vector $\mathcal{V} = (V, 1 - V)$, where V is the median of $2k+1$ independent, uniformly on $[0,1]$ distributed r.v.

Another example which fits not exactly in the model of a random split tree but is of similar type is the random recursive tree. The recursion for the path length X_n of the random recursive tree is of the slightly modified form

$$X_n \stackrel{\mathcal{D}}{=} X_K^{(1)} + X_{n-K}^{(2)} + K.$$

$(X_i^{(k)})$ are i.i.d. copies of X_i, $(X_i^{(1)}), (X_i^{(2)}), K$ are independent and K is uniformly distributed on $\{1, \ldots, n-1\}$. For this tree the limit law for X_n was proved by a similar method in Dobrow and Fill (1999). In this paper the authors also derive explicitly the higher moments of the limiting distribution in terms of the ζ-function.

Also quadtrees (discussed for its own in Neininger and Rüschendorf (1999)) and m-ary search trees are in the range of Theorem 2.2.3.

Quadtrees

The random d-dimensional quadtree (see section 1.4) is the random split tree with branch factor $b = 2^d$, vertex capacity $s_0 = 1$, distribution parameters $s = 1, s_1 = 0$ and splitting vector $\mathcal{V} = \langle U \rangle$. Here U is uniformly on $[0,1]^d$ distributed and $\langle U \rangle$ defined by (1.100). Therefore the splitter V is a product of d independent and uniformly on $[0,1]$ distributed r.v. This implies

$$\mu = b\,\mathbb{E}\left[V \ln(1/V)\right] = \frac{d}{2}. \tag{2.47}$$

The mean of the internal path length Y_n of a random d-dimensional quadtree has been found in Flajolet, Labelle, Laforest, and Salvy (1995):

$$\mathbb{E}\,Y_n = \frac{2}{d} n \ln n + c_d n + o(1). \tag{2.48}$$

(The first order asymptotic has been given before independently by Flajolet, Gonnet, Puech, and Robson (1993) and Devroye and Laforest (1990).) The conditions of Theorem 2.2.3 are satisfied. The scaled internal path length converges in ℓ_2 to the unique fixed point of

$$X \stackrel{\mathcal{D}}{=} \sum_{k=0}^{2^d-1} \langle U \rangle_k X^{(k)} + C(\langle U \rangle) \tag{2.49}$$

in $M_{0,2}$ where $X^{(k)}, U$ are independent, $X^{(k)} \sim X, U$ uniformly distributed on $[0,1]^d$ and the entropy functional given by

$$C(x) = 1 + \frac{2}{d} \sum_{k=0}^{2^d-1} x_k \ln x_k. \tag{2.50}$$

Part (b) of Theorem 2.2.3 gives the first order asymptotic of the variance of the internal path length

Corollary 2.3.1 (Variance for the internal path length in quadtrees) *The variance of the internal path length Y_n in a random d-dimensional quadtree satisfies*

$$Var(Y_n) \sim v_d n^2 \tag{2.51}$$

with

$$v_d = \frac{21 - 2\pi^2}{9d(1-(2/3)^d)}. \tag{2.52}$$

Proof: Let X denote the limit of the scaled path length. (2.49) and the independence properties imply

$$\mathbb{E}\, X^2 = \left(1 - \left(\frac{2}{3}\right)^d\right)^{-1} \mathbb{E}\, C^2(\langle U \rangle). \tag{2.53}$$

By calculation as in the proof of Lemma 2.2.2

$$\mathbb{E}\, C^2(\langle U \rangle) = -1 + \frac{4}{d^2} \mathbb{E} \left(\sum_{k=0}^{2^d-1} \langle U \rangle_k \ln \langle U \rangle_k \right)^2 \tag{2.54}$$

$$= -1 + \frac{4}{d^2} \sum_{i,j=0}^{2^d-1} \mathbb{E}\, [\langle U \rangle_i \langle U \rangle_j \ln \langle U \rangle_i \ln \langle U \rangle_j]. \tag{2.55}$$

The distribution of the factors $\langle U \rangle_i \langle U \rangle_j \ln \langle U \rangle_i \ln \langle U \rangle_j$ only depends on the number of digits in which the dual representations of i and j differ (see (1.96), (1.97)). Therefore

$$\mathbb{E} \left(\sum_{k=0}^{2^d-1} \langle U \rangle_k \ln \langle U \rangle_k \right)^2 = \sum_{h=0}^{d} 2^d \binom{d}{h} l_h. \tag{2.56}$$

l_h can be calculated by first applying the functional equation of the logarithm. This yields d^2 terms of the form

$$\int_{[0,1]^d} \prod_{i=1}^{d-h} u_i^2 \prod_{i=d-h+1}^{d} (u_i(1-u_i)) \ln \widetilde{u}_k \ln \widetilde{u}_l \, \mathrm{d}\lambda^d(u) \tag{2.57}$$

with $\widetilde{u}_k = u_k$ for $k \le d-h$ and $\widetilde{u}_k = 1 - u_k$ for $k > d-h$. Then distinguish the cases $1 \le k, l \le d-h$ and $d-h+1 \le k, l \le d$ for $k = l$ and $k \ne l$ and finally $1 \le k \le d-h < l \le d$. The arising integrals can be calculated explicitly. This implies the representation

$$v_d = \left(1 - \left(\frac{2}{3}\right)^d\right)^{-1} \left[-1 + \frac{4}{d^2}\left(\frac{2}{3}\right)^d \sum_{h=0}^{d} \binom{d}{h} \left(\frac{1}{2}\right)^h s_h\right] \tag{2.58}$$

where

$$s_h = \left(\frac{d}{3} - \frac{h}{2}\right)^2 + \frac{d}{9} + \left(\frac{5}{4} - \frac{\pi^2}{6}\right) h. \tag{2.59}$$

Now a simplification with the help of Maple[1] leads to the stated variance. ■

m-ary search trees

The random m-ary search tree (see Mahmoud (1992)) is the random split tree with branch factor $b = m$, vertex capacity $s_0 = m-1$, distribution parameters $s = m-1, s_1 = 0$ and split vector $\mathcal{V} = (V_1, \dots, V_m)$, where $V_1, \dots, V_m$ are the spacings of $m-1$ i.i.d. random variables uniformly distributed on $[0,1]$. For $U_1, \dots, U_{m-1}$ i.i.d. and uniformly distributed on $[0,1]$ denote by $U_{(1)}, \dots, U_{(m-1)}$ the order statistics of $U_1, \dots, U_{m-1}$. Then

$$\mathcal{V} \stackrel{\mathcal{D}}{=} (U_{(1)}, U_{(2)} - U_{(1)}, \dots, U_{(m-2)} - U_{(m-1)}, 1 - U_{(m-1)}). \tag{2.60}$$

The splitter V is given as the minimum of $m-1$ independent, uniformly on $[0,1]$ distributed r.v. This implies

$$\mu = b\,\mathbb{E}\left[V \ln(1/V)\right] = H_m - 1, \tag{2.61}$$

where H_m denotes the mth harmonic number, $H_m = \sum_{i=1}^m 1/i$. The mean of the internal path length Y_n for the random m-ary search tree has been analyzed in Mahmoud (1986):

$$\mathbb{E}\,Y_n = \frac{1}{H_m - 1} H_n(n+1) + c_m n + \mathcal{O}(n^\beta) \tag{2.62}$$

with $\beta < 1$. Substituting $H_n = \ln n + \gamma + O(1/n)$ in (2.62) with γ being Euler's constant $\mathbb{E}\,Y_n$ is of the form (2.28) with leading constant $\mu^{-1} = 1/(H_m - 1)$.

[1]This was done by P. Flajolet.

Thus the conditions of Theorem 2.2.3 are satisfied. The normalized path length converges in ℓ_2 to the unique fixed point of

$$X \stackrel{\mathcal{D}}{=} \sum_{k=1}^{m} V_k X^{(k)} + C(\mathcal{V}), \tag{2.63}$$

where $X^{(k)}, \mathcal{V}$ are independent, $X^{(k)} \sim X$ and $\mathcal{V} = (V_1, \ldots, V_m)$ is the vector of spacings of $m-1$ independent, uniformly on $[0,1]$ distributed r.v. The entropy functional is given here by

$$C(x) = 1 + \frac{1}{H_m - 1} \sum_{k=1}^{m} x_k \ln x_k. \tag{2.64}$$

In principle higher moments can be calculated from the fixed point equation (2.63). The first order asymptotic for the second order moment of the path length of m-ary search trees has already been achieved by generating function methods (cf. Mahmoud (1992, p. 142)).

Remarks: From the point of view of Theorem 2.2.3 it is a challenging task to identify those splitting vectors $\mathcal{V} = (V_1, \ldots, V_m)$, which induce an expansion (2.28) for the mean of the internal path length. For the case of the random quadtree and the m-ary search tree these expansions were derived by generating function analysis. A new and general approach to this problem was given in Rösler (1999) (see also Bruhn (1996)) using arguments from renewal theory. In particular Rösler derived the expansion for the median–of–$(2k+1)$ search tree *via* this new method. These techniques might be appropriate to characterize the splitting vectors satisfying the conditions of Theorem 2.2.3.

The analysis given in this chapter does not cover the cases of digital structures as tries or digital search trees, since Theorem 2.2.3 is restricted to continuous splitting vectors. In the case of digital structures this analysis leads to a degeneration of the limiting equation in the sense that the entropy functional C cancels out. Then the Dirac measure in 0 turns out to be the limit of the scaled path length. This indicates that the scaling of the internal path length by n^{-1} is of the wrong order of magnitude. In fact for the case of the digital search tree in the asymmetric Bernoulli model Jacquet and Szpankowski (1995) showed that the variance of the internal path length is of the order $n \ln n$ and that the (correctly) normalized path length tends to the standard normal distribution.

Finally we remark that the estimate for large deviations in part (d) of Theorem 2.2.3 has been improved by McDiarmid and Hayward (1996) for the case of the binary search tree and the median–of–$(2k+1)$ tree. A more general result in this direction for the binary search tree was given in Knessl and Szpankowski (1999).

Chapter 3

Multiple Quickselect

The *Find* algorithm (also called Quickselect or one-sided Quicksort) was introduced in Hoare (1961, 1962) for finding order statistics in a given list. Find is based on the same design principle as the related quicksort algorithm. The problem is to select the jth smallest element of an array containing n data. First, choose by some (randomized) rule a certain element of the array, the *pivot*, and bring it into its correct position. This means rearrange the array so that left of the pivot being only smaller, right to it being only greater elements. Now, if the position of the pivot is j we are done and return the pivot. If the pivot's order is greater than j, then the jth element must be in the part of the array with the smaller elements, otherwise in the part with the greater elements. Apply the procedure recursively to that segment which contains the desired statistic to continue the selection procedure and abandon the other subarray.

Multiple Quickselect is a variant of Find modified to search for more than one order statistic at a time. Multiple Quickselect works as Find bringing first the pivot to its right position. Since two or more statistics are sought, one or both of the generated subarrays might contain statistics to be reported. Thus multiple Quickselect may be applied recursively to one or both subarrays to search for the desired statistics smaller respectively greater than the pivot. For an exact description of the algorithm see Lent and Mahmoud (1996).

The running time of multiple Quickselect is measured by the number of key comparisons done during the execution of the algorithm. For the probabilistic analysis of the running time assume the uniform model, i.e. assume all permutations of the ranks of the data to be equally likely. Denote by $p \geq 1$ the number of order statistics sought. There are several reasonable models for these statistics. So far the orders of the statistics searched for have been assumed to be fixed, uniformly distributed over their range, increasing as a fixed fraction of the number of keys n, and also the number p of statistics itself has been considered to be increasing with n as a fixed fraction of n or to be of the form $n - l$ with a fixed $l \geq 1$. Let $C_n^{(m_1,\ldots,m_p)}$ denote the number of key comparisons of multiple

Quickselect in the uniform model seeking for fixed order statistics $m_1, \ldots, m_p$. Denote by $\overline{C}_p(n)$ the number of key comparisons of multiple Quickselect in the uniform model if the statistics are uniformly distributed over

$$\{(m_1, \ldots, m_p) \in \{1, \ldots, n\}^p \mid m_1 < m_2 < \ldots < m_p\}. \tag{3.1}$$

In the case $p = 1$, which is Find, Mahmoud, Moddares, and Smythe (1995) showed

$$\mathbb{E}\,\overline{C}_1(n) = 3n - 8H_n + 13 - 8H_n/n = 3n + o(n), \tag{3.2}$$

$$\mathrm{Var}(\overline{C}_1(n)) \sim n^2 \quad \text{for } n \to \infty, \tag{3.3}$$

and weak convergence for the normalized version of $\overline{C}_1(n)$. For the r.v. $C_n^{(m_1,\ldots,m_p)}$ with fixed orders $m_1, \ldots, m_p$ and $p = 1$ Knuth (1972) gave an exact formula for $\mathbb{E}\,C_n^{(m)}$, Kirschenhofer and Prodinger (1998) found an explicit formula for $\mathrm{Var}\ C_n^{(m)}$. Weak convergence for the normalized versions of $C_n^{(m)}$ was proved in Mahmoud et al. (1995). In the case $p = 1$ when m is a fixed fraction of n, i.e. $m \sim \varrho n$ for a $0 < \varrho < 1$ the formula

$$\mathbb{E}\,C_n^{(m)} = e_\varrho n + o(n) \text{ with } e_\varrho = 2 - 2(\varrho \ln \varrho + (1-\varrho)\ln(1-\varrho)) \tag{3.4}$$

can be deduced directly from Knuth's formula. $\mathrm{Var}\ C_n^{(m)}$ in this case was derived asymptotically in Paulsen (1997) and Kirschenhofer and Prodinger (1998). Paulsen also considered higher moments. Weak convergence for the scaled versions follows from the work of Grübel and Rösler (1996), where a limit law for the whole *Find process* $(C_n^{([\varrho n+1])})_{0\le\varrho\le1}$ in the space $D([0,1])$ of càdlàg functions on $[0,1]$ endowed with the Skorokhod topology is given (see also Grübel (1998)).

In the case $p \geq 1$ Lent and Mahmoud (1996) gave the asymptotic

$$\mathbb{E}\,\overline{C}_p(n) = (2H_p + 1)n - 8p \ln n + \mathcal{O}(1). \tag{3.5}$$

An explicit (non asymptotic) formula for $\mathbb{E}\,\overline{C}_p(n)$ was given in Prodinger (1995), where also the exact expansion for $\mathbb{E}\,C_n^{(m_1,\ldots,m_p)}$ is derived. The variance of $\overline{C}_p(n)$ for $p \geq 2$ was calculated exactly in Panholzer and Prodinger (1998). In this work also cases where p is a fraction of n or $p = n - l$ for a fixed $l \geq 1$ are considered as well as median–of–three variants of multiple quickselect. For the median–of–three variant the pivot is drawn as the median of three independent samples uniformly distributed over the keys. In the case $p = 1$ of Find for the median–of–three variant $\mathbb{E}\,C_n^{(m)}$ and $\mathbb{E}\,\overline{C}_p(n)$ have been given asymptotically in Kirschenhofer, Martínez, and Prodinger (1997), (see also Anderson and Brown (1992)). The model where $m \sim \varrho n$ is a fraction of n was treated for the median–of–$(2k+1)$ version of Find in Grübel (1999). The problem of finding an optimal

k for a median–of–$(2k+1)$ variant for Quicksort or Find has been investigated in Martínez and Roura (1998).

Also the passes of multiple Quickselect have been analyzed. These are the number of recursive calls of the algorithm during its execution. Denote by $P_n^{(m_1,\ldots,m_p)}$ and $\overline{P}_p(n)$ the number of passes in the uniform model with fixed order statistics $m_1 \ldots, m_p$ respectively uniformly distributed statistics. In the Find case $p=1$, $\mathbb{E}\,P_n^{(m)}$ has already been given exactly in Arora and Dent (1969). A formula for Var $P_n^{(m)}$ was derived in Kirschenhofer and Prodinger (1998). First moments for the median–of–three variant were treated in Kirschenhofer et al. (1997). For $p \geq 1$ Kirschenhofer and Prodinger (1998) calculated $\mathbb{E}\,P_n^{(m_1,\ldots,m_p)}$ and $\mathbb{E}\,\overline{P}_p(n)$. Second moments for $\overline{P}_p(n)$ and models with p being a fraction of n or of the form $p = n - l$ for fixed l were treated in Panholzer and Prodinger (1998).

In the first section of this chapter a limit law for $\overline{C}_p(n)$ in the median–of–three variant will be derived based on the contraction method. In the second section this limit law is specialized to the case $p=1$, which is the median–of–three version of Find. Further results on the asymptotics of all moments, estimates for large deviations and results concerning the Laplace and Fourier transform are given for the median–of–three Find.

Remark: Originally I treated a limit law for $\overline{C}_p(n)$ for the standard multiple Quickselect without median–of–three selection of the pivot element by means of the contraction method. This was found independently by the same approach in Mahmoud and Smythe (1998). Since the same method also works for the median–of–three selection this variant is treated here in order to keep originality.

3.1 Median–of–three multiple Quickselect

Denote by $\overline{C}_p(n)$ the number of key comparisons of multiple Quickselect applied to an array with n data. The orders of the data are assumed to be randomly permuted and the statistics sought are assumed to be uniformly distributed over the set in (3.1). The pivot is drawn as the median of three independent samples uniformly distributed over the data. Assume that the uniformity assumptions still hold in the subarrays after the pivot is inserted to its final position. This can be achieved using an appropriate procedure for the rearrangement of the array. Let Z_n denote the position of the pivot after the first partitioning step. Z_n is distributed as the median of three independent and uniformly on $\{1,\ldots,n\}$ distributed r.v. Let $M^{p,n} = (M_1^{p,n},\ldots,M_p^{p,n})$ denote the statistics sought, i.e. $M^{p,n}$ is uniformly distributed over the set in (3.1). For the insertion of the pivot element we need $n-1$ key comparisons. In the case $Z_n = M_i^{p,n}$ for some $i \in \{1,\ldots,p\}$ we have to select recursively $(i-1)$ statistics in $Z_n - 1$ keys and independently $(p-i)$ statistics in $n - Z_n$ keys. In the case $M_i^{p,n} < Z_n < M_{i+1}^{p,n}$ for a $i \in \{1, \ldots, p-1\}$ we have to select i statistics in $Z_n - 1$ keys and $(p-i)$ statistics in $n - Z_n$ keys. The cases $Z_n < M_1^{p,n}$ and $Z_n > M_p^{p,n}$ lead to only one recursive call of the algorithm. This leads to the following recursive distributional equation for $\overline{C}_p(n)$:

$$\begin{aligned}\overline{C}_p(n) \stackrel{\mathcal{D}}{=}\ & \mathbf{1}_{\{Z_n > M_p^{p,n}\}} \overline{C}_p^*(Z_n - 1) + \mathbf{1}_{\{Z_n < M_1^{p,n}\}} \overline{C}_p^{**}(n - Z_n) \\ & + \sum_{i=1}^{p-1} \mathbf{1}_{\{M_i^{p,n} < Z_n < M_{i+1}^{p,n}\}} (\overline{C}_i^*(Z_n - 1) + \overline{C}_{p-i}^{**}(n - Z_n)) \\ & + \sum_{i=1}^{p} \mathbf{1}_{\{Z_n = M_i^{p,n}\}} (\overline{C}_{i-1}^*(Z_n - 1) + \overline{C}_{p-i}^{**}(n - Z_n)) \\ & + n - 1, \end{aligned} \tag{3.6}$$

with $M^{p,n}$, Z_n, $\overline{C}_i^*(j)$, $\overline{C}_i^{**}(j)$, $1 \le i \le p, 1 \le j \le n-1$ being independent, $\overline{C}_i^*(j) \sim \overline{C}_i^{**}(j) \sim \overline{C}_i(j)$ and $M^{p,n}$, Z_n distributed as described before. We do not count the comparisons for finding the median. The first moment of $\overline{C}_p(n)$ is given in Panholzer and Prodinger (1998) (see also Panholzer (1997)):

$$\mathbb{E}\, \overline{C}_p(n) = c_p n + \mathcal{O}(\ln n) \tag{3.7}$$

with

$$c_p = \frac{12}{7} H_p + \frac{r(p)}{49(p+1)(p+2) \cdot \ldots \cdot (p+7)} \tag{3.8}$$

and

$$\begin{aligned} r(p) =\ & 37p^7 + 1036p^6 + 11914p^5 + 72520p^4 \\ & + 250453p^3 + 485884p^2 + 483516p + 246960. \end{aligned} \tag{3.9}$$

For the scaling we assume

$$\mathrm{Var}(\overline{C}_p(n)) \sim w_p n^2 \tag{3.10}$$

with some constant $w_p > 0$. This will be verified later. The normalized version

$$Y_n^{(p)} := \frac{\overline{C}_p(n) - \mathbb{E}\, \overline{C}_p(n)}{n} \tag{3.11}$$

by a straightforward calculation satisfies

$$\begin{aligned}
Y_n^{(p)} \stackrel{\mathcal{D}}{=} \ & \mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} \left(Y_{Z_n-1}^{(p)} + c_p \right) \\
& + \mathbf{1}_{\{Z_n < M_1^{p,n}\}} \frac{n - Z_n}{n} \left(\overline{Y}_{n-Z_n}^{(p)} + c_p \right) \\
& + \sum_{i=1}^{p-1} \mathbf{1}_{\{M_i^{p,n} < Z_n < M_{i+1}^{p,n}\}} \Bigg(\frac{Z_n - 1}{n} \left(Y_{Z_n-1}^{(i)} + c_i \right) \\
& \qquad + \frac{n - Z_n}{n} \left(\overline{Y}_{n-Z_n}^{(p-i)} + c_{p-i} \right) \Bigg) \\
& + \sum_{i=1}^{p} \mathbf{1}_{\{Z_n = M_i^{p,n}\}} \Bigg(\frac{Z_n - 1}{n} \left(Y_{Z_n-1}^{(i-1)} + c_{i-1} \right) \\
& \qquad + \frac{n - Z_n}{n} \left(\overline{Y}_{n-Z_n}^{(p-i)} + c_{p-i} \right) \Bigg) \\
& - c_p + \frac{n-1}{n} + R_n^{(p)},
\end{aligned} \tag{3.12}$$

with independencies and distributions analogously to (3.6). The $R_n^{(p)}$ depend on randomness and converge uniformly to zero. Since Z_n is the median of three independent, uniformly on $\{1, \dots, n\}$ distributed r.v. and $M^{p,n}$ is uniformly distributed over the set in (3.1), independent of Z_n, we derive for the scaled versions

$$\frac{1}{n}(Z_n, M^{p,n}) \xrightarrow{\mathbb{P}} (T, U_{(1)}, \dots, U_{(p)}) \tag{3.13}$$

where $T, U_{(1)}, \dots, U_{(p)}$ are independent, $U_1, \dots, U_p$ are uniformly on $[0,1]$ distributed and T is distributed as the median of three independent, uniformly on $[0,1]$ distributed r.v., i.e. beta$(2,2)$ distributed.

Lemma 3.1.1 *Let $T, U_1, \dots, U_p$ be independent, $U_1, \dots, U_p$ uniformly on $[0,1]$ distributed and $T \sim$ beta$(2,2)$, then*

$$\begin{aligned}
& \mathbb{E}\Bigg[\left(\mathbf{1}_{\{T > U_{(p)}\}} T + \mathbf{1}_{\{T < U_{(1)}\}} (1 - T) \right) c_p \\
& \qquad + \sum_{i=1}^{p-1} \mathbf{1}_{\{U_{(i)} < T < U_{(i+1)}\}} \left(T c_i + (1-T) c_{p-i} \right) \Bigg] \\
& = c_p - 1.
\end{aligned} \tag{3.14}$$

Proof: The r.v. $Y_j^{(i)}, \overline{Y}_j^{(i)}$ in the modified recursion (3.12) are centered and independent of everything else. This implies

$$\begin{aligned}\mathbb{E}\Bigg[&\left(\mathbf{1}_{\{Z_n>M_p^{p,n}\}}\frac{Z_n-1}{n}+\mathbf{1}_{\{Z_n<M_1^{p,n}\}}\frac{n-Z_n}{n}\right)c_p\\&+\sum_{i=1}^{p-1}\mathbf{1}_{\{M_i^{p,n}<Z_n<M_{i+1}^{p,n}\}}\left(\frac{Z_n-1}{n}c_i+\frac{n-Z_n}{n}c_{p-i}\right)\\&+\sum_{i=1}^{p}\mathbf{1}_{\{Z_n=M_i^{p,n}\}}\left(\frac{Z_n-1}{n}c_{i-1}+\frac{n-Z_n}{n}c_{p-i}\right)\\&-c_p+\frac{n-1}{n}+R_n^{(p)}\Bigg]\\&=0. \qquad (3.15)\end{aligned}$$

All the quantities are bounded. By (3.13) we can pass to the limit. This leads to the assertion. ■

Now assume that for indices $i<p$ convergence in distribution for $Y_n^{(i)}$ to a $Y^{(i)}$ is already shown. Then (3.12) and the convergence in (3.13) suggest that a limit $Y^{(p)}$ of $Y_n^{(p)}$ should satisfy the limiting equation

$$\begin{aligned}Y^{(p)}\stackrel{\mathcal{D}}{=}\ &\mathbf{1}_{\{T>U_{(p)}\}}T\left(Y^{(p)}+c_p\right)+\mathbf{1}_{\{T<U_{(1)}\}}(1-T)\left(\overline{Y}^{(p)}+c_p\right)\\&+\sum_{i=1}^{p-1}\mathbf{1}_{\{U_{(i)}<T<U_{(i+1)}\}}\Big(T\big(Y^{(i)}+c_i\big)+(1-T)\Big(\overline{Y}^{(p-i)}+c_{p-i}\Big)\Big)\\&-c_p+1, \qquad (3.16)\end{aligned}$$

where $T,\ U_1,\ \ldots,\ U_p,\ Y^{(1)},\ \overline{Y}^{(1)},\ \ldots,Y^{(p)},\ \overline{Y}^{(p)}$ are independent $U_1,\ldots,U_p$ uniformly distributed on $[0,1]$, $T\sim\text{beta}(2,2)$ and $Y^{(i)},\ \overline{Y}^{(i)}$ are distributed as the weak limits of $(Y_n^{(i)})$ and $Y^{(p)}\sim\overline{Y}^{(p)}$.

Now we define successively operators

$$S_1,\ldots,S_r : M^1(\mathbb{R},\mathcal{B})\to M^1(\mathbb{R},\mathcal{B}) \qquad (3.17)$$

which are contractions on $(M_{0,2},\ell_2)$ where the fixed points of S_j in $M_{0,2}$ for $j<r$ are used for the definition of S_r. For $r=1$ define $S_1 : M^1(\mathbb{R},\mathcal{B})\to M^1(\mathbb{R},\mathcal{B})$ by

$$S_1(\mu):\stackrel{\mathcal{D}}{=}\mathbf{1}_{\{T>U\}}T(Z+c_1)+\mathbf{1}_{\{T<U\}}(1-T)(\overline{Z}+c_1)-c_1+1, \qquad (3.18)$$

with $T,\ U,\ Z,\ \overline{Z}$ being independent, U uniformly on $[0,1]$ distributed, $T \sim$ beta$(2,2)$ and $Z \sim \overline{Z} \sim \mu$. The contraction property of S_1 on $(M_{0,2}, \ell_2)$ can be deduced as in Lemma 3.1.2 below. Now assume operators $S_j : M^1(\mathbb{R}, \mathcal{B}) \to M^1(\mathbb{R}, \mathcal{B})$ with the contraction property on $(M_{0,2}, \ell_2)$ are already defined for $j < r$. Then define $S_r : M^1(\mathbb{R}, \mathcal{B}) \to M^1(\mathbb{R}, \mathcal{B})$ by

$$\begin{aligned} S_r(\mu) :\stackrel{\mathcal{D}}{=} &\mathbf{1}_{\{T>U_{(r)}\}} T(Z + c_r) + \mathbf{1}_{\{T<U_{(1)}\}}(1-T)(\overline{Z} + c_r) \\ &+ \sum_{i=1}^{r-1} \mathbf{1}_{\{U_{(i)}<T<U_{(i+1)}\}} \Big(T\big(Y^{(i)} + c_i\big) + (1-T)\Big(\overline{Y}^{(r-i)} + c_{p-i}\Big) \Big) \\ &- c_r + 1 \end{aligned} \tag{3.19}$$

where $T,\ U_1, \dots, U_r,\ Z,\ \overline{Z},\ Y^{(1)},\ \overline{Y}^{(1)}, \dots, Y^{(r-1)},\ \overline{Y}^{(r-1)}$ are independent, $U_1, \dots, U_r$ are uniformly distributed on $[0,1]$, $T \sim$ beta$(2,2)$, $Z \sim \overline{Z} \sim \mu$ and $Y^{(i)}, \overline{Y}^{(i)}$ being versions of the fixed point of S_i in $M_{0,2}$ for $i = 1, \dots, r-1$.

Lemma 3.1.2 *$S_r : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. ℓ_2:*

$$\ell_2(S_r(\mu), S_r(\nu)) \leq \sqrt{3/5}\, \ell_2(\mu, \nu) \quad \textit{for all} \quad \mu, \nu \in M_{0,2}. \tag{3.20}$$

Proof: Obviously $\mathrm{Var}(S_r(\mu)) < \infty$ and $\mathbb{E}\, S_r(\mu) = 0$ from Lemma 3.1.1 for all $\mu \in M_{0,2}$. So $S_r : M_{0,2} \to M_{0,2}$ is well defined. Let $\mu, \nu \in M_{0,2}$ and choose independent $T,\ U_1, \dots, U_r,\ (V, W),\ (\overline{V}, \overline{W}),\ Y^{(1)}, \overline{Y}^{(1)}, \dots, Y^{(r-1)}, \overline{Y}^{(r-1)}$, where $T, U_1, \dots, U_r$ are as before, $Y^{(i)}, \overline{Y}^{(i)}$ fixed points of S_i for $i = 1, \dots, r-1$ and $V \sim \overline{V} \sim \mu$, $W \sim \overline{W} \sim \nu$ optimal ℓ_2-couplings of μ, ν, then

$$\begin{aligned} S_r(\mu) \stackrel{\mathcal{D}}{=} &\mathbf{1}_{\{T>U_{(r)}\}} T(V + c_r) + \mathbf{1}_{\{T<U_{(1)}\}}(1-T)(\overline{V} + c_r) \\ &+ \sum_{i=1}^{r-1} \mathbf{1}_{\{U_{(i)}<T<U_{(i+1)}\}} \Big(T\big(Y^{(i)} + c_i\big) + (1-T)\Big(\overline{Y}^{(r-i)} + c_{p-i}\Big) \Big) \\ &- c_r + 1, \end{aligned} \tag{3.21}$$

$$\begin{aligned} S_r(\nu) \stackrel{\mathcal{D}}{=} &\mathbf{1}_{\{T>U_{(r)}\}} T(W + c_r) + \mathbf{1}_{\{T<U_{(1)}\}}(1-T)(\overline{W} + c_r) \\ &+ \sum_{i=1}^{r-1} \mathbf{1}_{\{U_{(i)}<T<U_{(i+1)}\}} \Big(T\big(Y^{(i)} + c_i\big) + (1-T)\Big(\overline{Y}^{(r-i)} + c_{p-i}\Big) \Big) \\ &- c_r + 1. \end{aligned} \tag{3.22}$$

With independence and the centered mean properties we derive

$$\begin{aligned}
\ell_2^2(S_r(\mu), S_r(\nu)) &\leq \mathbb{E}\left(\mathbf{1}_{\{T>U_{(r)}\}}T(V-W) + \mathbf{1}_{\{T<U_{(1)}\}}(1-T)(\overline{V}-\overline{W})\right)^2 \\
&= \mathbb{E}\left[\mathbf{1}_{\{T>U_{(r)}\}}T^2(V-W)^2 + \mathbf{1}_{\{T<U_{(1)}\}}(1-T)^2(\overline{V}-\overline{W})^2\right] \\
&\leq \mathbb{E}\,[T^2+(1-T)^2]\ell_2^2(\mu,\nu) \\
&= \frac{3}{5}\,\ell_2^2(\mu,\nu). \qquad (3.23)
\end{aligned}$$

This Lipschitz constant can be improved considering also the indicators in (3.23). ■

Theorem 3.1.3 (Limit theorem for the running time of multiple Quickselect with median–of–three partitioning) *The normalized number of key comparisons $Y_n^{(p)}$ of multiple Quickselect with uniformly distributed statistics sought and median–of–three partitioning converges in the ℓ_2 metric to the unique fixed point $Y^{(p)}$ in $M_{0,2}$ of the limiting operator S_p given in (3.19),*

$$\ell_2(Y_n^{(p)}, Y^{(p)}) \to 0 \quad \textit{for} \quad n \to \infty. \qquad (3.24)$$

Proof: The theorem is proved by induction on the number p of statistics sought. In the Find case $p=1$ Mahmoud, Modarres and Smythe (1995) showed $\ell_2(Y_n^{(1)}, Y^{(1)}) \to 0$ for the Find algorithm where the pivot is chosen uniformly over the array of data. This proof directly extends to the case of median–of–three partitioning. So the assertion is true for $p=1$. For the induction step $p-1 \to p$ assume

$$\ell_2(Y_n^{(i)}, Y^{(i)}) \to 0 \quad \text{for} \quad n \to \infty \quad \text{and} \quad i = 1, \dots, p-1. \qquad (3.25)$$

Write (3.12) as

$$Y_n^{(p)} \stackrel{\mathcal{D}}{=} A_n + B_n + \sum_{i=1}^{p-1} C_n^{(i)} + \sum_{i=1}^{p} D_n^{(i)} + E_n \qquad (3.26)$$

with

$$A_n := \mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n}\left(Y_{Z_n-1}^{(p)} + c_p\right), \qquad (3.27)$$

$$B_n := \mathbf{1}_{\{Z_n < M_1^{p,n}\}} \frac{n - Z_n}{n}\left(\overline{Y}_{n-Z_n}^{(p)} + c_p\right), \qquad (3.28)$$

$$C_n^{(i)} := \mathbf{1}_{\{M_i^{p,n}<Z_n<M_{i+1}^{p,n}\}}\left(\frac{Z_n-1}{n}\left(Y_{Z_n-1}^{(i)}+c_i\right)\right.$$
$$\left.+\frac{n-Z_n}{n}\left(\overline{Y}_{n-Z_n}^{(p-i)}+c_{p-i}\right)\right), \tag{3.29}$$

$$D_n^{(i)} := \mathbf{1}_{\{Z_n=M_i^{p,n}\}}\left(\frac{Z_n-1}{n}\left(Y_{Z_n-1}^{(i-1)}+c_{i-1}\right)\right.$$
$$\left.+\frac{n-Z_n}{n}\left(\overline{Y}_{n-Z_n}^{(p-i)}+c_{p-i}\right)\right), \tag{3.30}$$

$$E_n := c_p+\frac{n-1}{n}+R_n^{(p)}. \tag{3.31}$$

The independencies and distributions are as in (3.12). For the fixed point $Y^{(p)}$ of S_p in $M_{0,2}$ we have the representation

$$Y^{(p)} \stackrel{\mathcal{D}}{=} A+B+\sum_{i=1}^{p-1}C^{(i)}+E \tag{3.32}$$

with

$$A := \mathbf{1}_{\{T>U_{(p)}\}}T\left(Y^{(p)}+c_p\right), \tag{3.33}$$

$$B := \mathbf{1}_{\{T<U_{(1)}\}}(1–T)\left(\overline{Y}^{(p)}+c_p\right), \tag{3.34}$$

$$C^{(i)} := \mathbf{1}_{\{U_{(i)}<T<U_{(i+1)}\}}\left(T\left(Y^{(i)}+c_i\right)+(1–T)\left(\overline{Y}^{(p-i)}+c_{p-i}\right)\right), \tag{3.35}$$

$$E := -c_p+1. \tag{3.36}$$

where $Y^{(1)},\overline{Y}^{(1)},\dots,Y^{(p)},\overline{Y}^{(p)}$ are as in (3.16). Furthermore assume $(Y^{(i)},Y_n^{(i)}),(\overline{Y}^{(i)},\overline{Y}_n^{(i)})$ to be optimal ℓ_2-couplings for all $n\in\mathbb{N}$ and $1\le i\le p$. Then it follows

$$\begin{aligned}
&\ell_2^2(Y_n^{(p)},Y^{(p)})\\
&\quad= \ell_2^2\left(A_n+B_n+\sum_{i=1}^{p-1}C_n^{(i)}+\sum_{i=1}^{p}D_n^{(i)}+E_n, A+B+\sum_{i=1}^{p-1}C^{(i)}+D\right)\\
&\quad\le \mathbb{E}\left((A_n-A)+(B_n-B)+\sum(C_n^{(i)}-C^{(i)})+\sum_{i=1}^{p}D_n^{(i)}+(E_n-E)\right)^2\\
&\quad= \mathbb{E}\left[(A_n-A)^2+(B_n-B)^2+\sum(C_n^{(i)}-C^{(i)})^2\right]+o(1).
\end{aligned} \tag{3.37}$$

The mixed terms are zero or $o(1)$ by independence, the zero mean properties, bounded norms resulting from the induction hypothesis and (3.13). Furthermore, $\mathbb{E}\,[\sum (D_n^{(i)})^2]$ and $\mathbb{E}\,(E_n - E)^2$ are converging to zero. It is

$$\begin{aligned}
&\mathbb{E}\,(A_n - A)^2 \\
&\quad = \mathbb{E}\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n}\left(Y_{Z_n-1}^{(p)} + c_p\right) - \mathbf{1}_{\{T > U_{(p)}\}} T\left(Y^{(p)} + c_p\right)\right)^2 \\
&\quad = \mathbb{E}\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n}\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)\right. \\
&\qquad\qquad \left. + \left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} - \mathbf{1}_{\{T > U_{(p)}\}} T\right)\left(Y^{(p)} + c_p\right)\right)^2 \\
&\quad = \mathbb{E}\left[\mathbf{1}_{\{Z_n > M_p^{p,n}\}}\left(\frac{Z_n - 1}{n}\right)^2\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)^2\right. \\
&\qquad\qquad + \left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} - \mathbf{1}_{\{T > U_{(p)}\}} T\right)^2\left(Y^{(p)} + c_p\right)^2 \\
&\qquad\qquad + \ 2\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n}\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)\right) \\
&\qquad\qquad\qquad \left. \times\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} - \mathbf{1}_{\{T > U_{(p)}\}} T\right)\left(Y^{(p)} + c_p\right)\right]. \qquad (3.38)
\end{aligned}$$

From (3.13) it follows

$$\mathbb{E}\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} - \mathbf{1}_{\{T > U_{(p)}\}} T\right)^2 \to 0 \quad \text{for} \quad n \to \infty. \qquad (3.39)$$

Therefore the second summand in (3.38) converges to zero. With the Cauchy-Schwarz inequality and (3.39) the third summand in its absolute value is estimated from above by

$$\begin{aligned}
&2\,\mathbb{E}\left[\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}} \frac{Z_n - 1}{n} - \mathbf{1}_{\{T > U_{(p)}\}} T\right)^2\left(Y^{(p)} + c_p\right)^2\right]^{1/2} \\
&\qquad\qquad\qquad \times\,\mathbb{E}\left[\left(Y_{Z_{n-1}}^{(p)} - Y^{(p)}\right)^2\right]^{1/2} \\
&\quad = o(1)\,\mathbb{E}\left[\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)^2\right]^{1/2} \\
&\quad \leq o(1)\,\mathbb{E}\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)^2 + o(1). \qquad (3.40)
\end{aligned}$$

The last inequality holds since both sides are $o(1)$ if the expectation is less than 1. This implies

$$\begin{aligned}
&\mathbb{E}\,(A_n - A)^2 \\
&= \mathbb{E}\left[\left(\mathbf{1}_{\{Z_n > M_p^{p,n}\}}\left(\frac{Z_n - 1}{n}\right)^2 + o(1)\right)\left(Y_{Z_n-1}^{(p)} - Y^{(p)}\right)^2\right] + o(1) \\
&= \mathbb{E}\left[\sum_{j=1}^{n} \mathbf{1}_{\{Z_n = j\}} \mathbf{1}_{\{j > M_p^{p,n}\}}\left(\left(\frac{j-1}{n}\right)^2 + o(1)\right)\left(Y_{j-1}^{(p)} - Y^{(p)}\right)^2\right] + o(1) \\
&\leq \sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\left(\frac{j}{n}\right)^2 + o(1)\right)\ell_2^2(Y_{j-1}^{(p)}, Y^{(p)}) + o(1) \qquad (3.41)
\end{aligned}$$

Analogously we derive

$$\begin{aligned}
&\mathbb{E}\,(B_n - B)^2 \\
&\leq \sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right)\ \ell_2^2(Y_j^{(p)}, Y^{(p)}) + o(1), \qquad (3.42)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\,(C_n^{(i)} - C^{(i)})^2 \\
&\leq 2\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right)\ \ell_2^2(Y_j^{(i)}, Y^{(i)}) + o(1). \qquad (3.43)
\end{aligned}$$

Denote by

$$s_j := \sum_{i=1}^{p-1} \ell_2^2(Y_j^{(i)}, Y^{(i)}) \text{ für } j \in \mathbb{N}. \qquad (3.44)$$

By the induction hypothesis the sequence (s_j) converges to zero. Altogether we derive

$$\begin{aligned}
\ell_2^2(Y_n^{(p)}, Y^{(p)}) \;\leq\; & 2\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right)\ \ell_2^2(Y_j^{(p)}, Y^{(p)}) \\
&+ 2\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right) s_j + o(1). \qquad (3.45)
\end{aligned}$$

It is

$$\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right) s_j \leq \sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1) s_j \qquad (3.46)$$

for n sufficiently large. This sum is converging so zero for $n \to \infty$: Let $\varepsilon > 0$. It exists a $n_0 \in \mathbb{N}$ with $s_j \leq \varepsilon/2$ for all $n \geq n_0$. Obviously $\mathbb{P}(\{Z_n \leq n_0\}) \to 0$ for $n \to \infty$. Choose $n_1 \in \mathbb{N}$ with $\mathbb{P}(\{Z_n \leq n_0\}) < \varepsilon/(2\max\{s_1, \ldots, s_{n_0}\})$ for all $n \leq n_1$. Then for $n \geq \max\{n_0, n_1\}$ it follows

$$\begin{aligned}
&\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)s_j \\
&\leq \sum_{j=1}^{n_0-1} \mathbb{P}(Z_n = j+1)s_j + \sum_{j=n_0}^{n-1} \mathbb{P}(Z_n = j+1)s_j \\
&\leq \max\{s_1, \ldots, s_{n_0}\}\,\mathbb{P}(Z_n \leq n_0) + \frac{\varepsilon}{2}\,\mathbb{P}(Z_n \geq n_0+1) \\
&< \varepsilon. \qquad (3.47)
\end{aligned}$$

This implies the recursion

$$\begin{aligned}
&\ell_2^2(Y_n^{(p)}, Y^{(p)}) \\
&\leq 2\sum_{j=1}^{n-1} \mathbb{P}(Z_n = j+1)\left(\frac{j^2}{n^2} + o(1)\right)\ \ell_2^2(Y_j^{(p)}, Y^{(p)}) + o(1). \qquad (3.48)
\end{aligned}$$

We can now conclude as in Theorem 1.1.2 or Theorem 2.2.3. This yields

$$\ell_2^2(Y_n^{(p)}, Y^{(p)}) \to 0 \quad \text{for} \quad n \to \infty. \qquad (3.49)$$

■

In particular Theorem 3.1.3 leads to the first order asymptotic of the variance of $\overline{C}_p(n)$:

$$\mathrm{Var}(\overline{C}_p(n)) \sim w_p n^2 \qquad (3.50)$$

with some $w_p > 0$. For the case of multiple Quickselect without median–of–three partitioning the leading constant in the corresponding expansion has been calculated explicitly in Mahmoud and Smythe (1998, Theorem 2), where also further properties of the corresponding limit distribution are stated.

3.2 Median–of–three Find

Consider the special case $p = 1$ of Find with median–of–three partitioning. The fixed point equation for the limit of the scaled version of $\overline{C}_1(n)$ is given by (3.18)

$$Y^{(1)} \stackrel{\mathcal{D}}{=} \mathbf{1}_{\{T>U\}}T(Y^{(1)} + c_1) + \mathbf{1}_{\{T<U\}}(1–T)(\overline{Y}^{(1)} + c_1) - c_1 + 1. \qquad (3.51)$$

The constant $c_1 = 5/2$ given by (3.8) has been calculated first in Kirschenhofer, Martínez, and Prodinger (1997). Equation (3.51) can be simplified to (cf. the remarks after (1.120))

$$Y^{(1)} \stackrel{\mathcal{D}}{=} X\left(Y^{(1)} + \frac{5}{2}\right) - \frac{3}{2}, \tag{3.52}$$

where $X, Y^{(1)}$ are independent and X has the Lebesgue-density

$$\frac{\mathrm{d}\mathbb{P}^X}{\mathrm{d}\lambda^1}(t) = 12t^2(1-t) \quad \text{for} \quad t \in [0,1]. \tag{3.53}$$

Thus $Y := Y^{(1)} + 5/2$ is the unique solution of

$$Y \stackrel{\mathcal{D}}{=} XY + 1 \tag{3.54}$$

in $M_{5/2,2}$, again X, Y being independent and X with density (3.53). In particular

$$\frac{\overline{C}_1(n)}{n} \xrightarrow{\mathcal{D}} Y \quad \text{for} \quad n \to \infty. \tag{3.55}$$

The higher order moments of Y can be calculated directly using (3.54):

$$\mathbb{E}\, Y^k = \sum_{j=0}^{k} \binom{k}{j} \mathbb{E}\, X^j \, \mathbb{E}\, Y^j. \tag{3.56}$$

It is

$$\mathbb{E}\, X^j = \frac{12}{(j+3)(j+4)} \quad \text{for} \quad j \geq 0. \tag{3.57}$$

With $m_k := \mathbb{E}\, Y^k$ for $k \geq 0$ it follows $m_0 = 1, m_1 = 5/2$ and using (3.56)

$$m_k = \left(12 + \frac{144}{k^2 + 7k}\right) \sum_{j=0}^{k-1} \binom{k}{j} \frac{m_j}{(j+3)(j+4)} \quad \text{for} \quad k \geq 2. \tag{3.58}$$

Existence and convergence of the Laplace transform of $\overline{C}_1(n)/n$ can be deduced similarly to Theorem 1.5.1 based on Lemma 4.1 and Theorem 4.2 in Rösler (1992). This implies convergence of all moments of $\overline{C}_1(n)/n$ to the corresponding moments of Y, especially

$$\mathbb{E}\,(\overline{C}_1(n))^k \sim m_k n^k \quad \text{for} \quad n \to \infty \tag{3.59}$$

with (m_k) given by (3.58). Furthermore analogously to (1.159) an estimate for large deviations can be established: For all $\lambda > 0$ there exists a $c_\lambda > 0$ so that for any sequence (a_n) of positive, real numbers holds:

$$\mathbb{P}(C_n \geq a_n) \leq c_\lambda \exp\left(-\lambda \frac{a_n}{n}\right). \tag{3.60}$$

The Fourier transform of Y is of the form

$$\mathbb{E}\exp(itY) = e^{it}\phi(t), \tag{3.61}$$

where $\phi : \mathbb{R} \to \mathbb{C}$ is smooth and given by the linear homogeneous differential equation of second order

$$t^2\phi''(t) + 8t\phi'(t) - 12(e^{it} - 1)\phi(t) = 0 \tag{3.62}$$

with the initial conditions $\phi(0) = 1$ and $\phi'(0) = 3i/2$. This can be computed following the line of Lemma 2 in Mahmoud, Moddares, and Smythe (1995), taking an additional second derivative.

Chapter 4

Interval splitting

Studies on the topic of interval splitting may be regarded as developing from a problem suggested to Kakutani by the physicist Araki (Kakutani (1975)). Let (X_n) denote a sequence of r.v. where X_1 is uniformly distributed on $[0, 1]$ and for $n \geq 2$ the conditional distribution of X_n given $X_1, \ldots, X_{n-1}$ is uniform on the largest of the n subintervals into which $[0, 1]$ is subdivided by $X_1, \ldots, X_{n-1}$. This model is called the Kakutani–model (or K–model) in comparison to the uniform model (or U–model) where a sequence of independent, uniformly on $[0, 1]$ distributed r.v. is considered. Kakutani (1975) proved that the empirical distribution of a related, more deterministic, sequence tends to the uniform distribution on $[0, 1]$ and conjectured that this should also hold almost surely for the K–model. For the early history of this problem and an alternative proof of Kakutani's result see Adler and Flatto (1977).

Kakutani's conjecture that the Glivenko-Cantelli result also holds for the K–model has been proved by van Zwet (1978) and independently by Slud (1978). Slud's method also extends to the generalization in which the dividing measure of the longest interval is arbitrary and not necessarily uniform. In addition, rates of convergence are considered by Slud.

Pyke (1980) based on the ideas of van Zwet investigated the convergence of the normalized spacings in the K–model in comparison to earlier results for the U–model.

In the uniform model let $I_k(n)$ denote the kth largest of the n intervals generated by $X_1, \ldots, X_{n-1}$ in $[0, 1]$. Bruss, Jammalamadaka, and Zhou (1990) first examined the problem of which sequences $(I_{k_n}(n))$ get covered infinitely often in the sense that the event $\{X_n \in I_{k_n}(n) \text{ infinitely often}\}$ has probability 1. Alsmeyer (1991) proved a zero-one law for this event and gave a condition on (k_n) to distinguish between the two cases. Mountford and Port (1993) rederived Alsmeyer's condition, investigated conditions for weak and strong laws of large numbers for the length of $(I_{k_n}(n))$ and gave a central limit theorem for scaled versions of $N_n := \text{card}\{1 \leq i \leq n : X_i \in I_{k_i}(i)\}$ in the case when $N_n \uparrow \infty$.

4.1 Random nested intervals

The investigation of interval splitting schemes of another type started with the papers of Chen, Goodman, and Zame (1984) and Chen, Lin, and Zame (1981). A uniformly on $[0,1]$ distributed r.v. U_0 partitions $[A_0, B_0] := [0,1]$ into two subintervals. Choose with probability $0 \leq p \leq 1$ the longer of these subintervals, otherwise the shorter one and denote it by $[A_1, B_1]$. Then choose U_1 uniformly on $[A_1, B_1]$ and iterate the procedure to define $[A_2, B_2]$. Further iteration gives a sequence of random nested intervals $([A_n, B_n])$ which shrinks to a limit point Y_p almost surely. Now the problem is to determine the distribution of Y_p. In the papers mentioned it is proved that Y_p has a beta(2,2) distribution if $p = 1$, i.e. choosing always the longer of the two intervals. In the case $p = 1/2$ the distribution of Y_p is the arcsine distribution which is the beta(1/2,1/2) distribution. Devroye, Letac, and Seshadri (1986) rediscovered these results and showed that Y_p has some beta distribution if and only if $p \in \{1/2, 1\}$.

Kennedy (1988) considered a related splitting scheme. Again start with the interval $[A_0, B_0] := [0,1]$. Let $[A_n, B_n]$ be already defined. Then let C_n, D_n be the minimum and maximum of k independent, uniformly on $[A_n, B_n]$ distributed r.v. and choose $[A_{n+1}, B_{n+1}]$ to be $[C_n, B_n], [A_n, D_n]$ or $[C_n, D_n]$ with probabilities p, q, r respectively, $p + q + r = 1$. This interval splitting scheme was motivated by the analysis of a randomized algorithm to locate local maxima of an arbitrary function. Kennedy showed that his splitting scheme shrinks to a point with beta$(k(p + r), k(q + r))$ distribution.

All these identifications of the limits of splitting schemes as well known distributions are based on certain invariance properties of the limits combined with the study of the moments. A general *moment method* for the characterization of distributions of a related type was discussed in Volodin, Kotz, and Johnson (1993). Applications of this moment method to the interval splitting schemes described above have been given in Johnson and Kotz (1990, 1995).

In the special case $k = 2, r = 1$ of Kennedy's scheme there is always the interval between two samples on $[A_n, B_n]$ chosen to be $[A_{n+1}, B_{n+1}]$. A generalization of this case has already been treated before by van Assche (1986). In this work the two samples are drawn from some joint distribution, where no restrictions on independence or uniformity of the samples are made. Van Assche states the problem in terms of products of two-dimensional random stochastic matrices. A stopping time which counts the number of steps until the length of the interval is less that a $t > 0$ is introduced in order to give a rate of convergence. This stopping time also measures the rate of convergence of the products of the related random stochastic matrices.

Another generalization of the case $k = 2, r = 1$ to higher dimensions has been treated recently by Letac and Scarsini (1998). In this work a tetrahedron T_0 in $\mathbb{R}^d$ which is given as the convex hull of $(d + 1)$ affinely independent points

$\alpha_1, \ldots, \alpha_{d+1} \in \mathbb{R}^d$ is considered. Then choose $(d+1)$ points independently and uniformly in T_0. The convex hull of these points defines a new tetrahedron T_1 contained in T_0. Iteration of this procedure gives a nested sequence of tetrahedra (T_n) which almost surely shrinks to a point Y. It is proved that the barycentric coordinates of Y with respect to $\alpha_1, \ldots, \alpha_{d+1}$ are Dirichlet distributed with parameter $(d+1, \ldots, d+1)$. Also this theorem is an application of results on products of random stochastic matrices given in Chamayou and Letac (1991, 1994). Letac and Scarsini also consider a rate of convergence in terms of the speed of shrinking of the tetrahedra.

Here, a new point of view of interval splitting schemes is presented. We derive rates of convergence in terms of related distributions which approximate the distribution of the point to which the splitting scheme shrinks.

4.2 Rate of convergence

We consider a splitting scheme which is a generalization of Kennedy's case $k = 2, r = 1$, and a special case of van Assche (1986). For $a, b \in \mathbb{R}$ define the affine map $A_{a,b}$ by

$$\begin{aligned} A_{a,b} : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto a + (b-a)x. \end{aligned} \tag{4.1}$$

Let $\mu \in M^1([0,1], \mathcal{B}^1)$ be a probability measure on $[0,1]$. We say that an r.v. is *chosen on* $[a,b]$ *according to* μ if its distribution is $A_{a,b}(\mu)$. In this section $[a,b]$ denotes the line segment with the end points a and b; we do not require $a \le b$ in this notations as for intervals. Let $[A_0, B_0] := [0,1]$. Assume that $[A_n, B_n]$ is already defined. Then choose $W_1^{(n)}, W_2^{(n)}$ independently on $[A_n, B_n]$ according to μ. Define $[A_{n+1}, B_{n+1}] := [W_1^{(n)}, W_2^{(n)}]$. If μ is the uniform distribution on $[0,1]$ this is Kennedy's scheme. In the following we will only assume $\mathbb{E}\,\mu = 1/2$. For the definition of the $W_i^{(n)}$ we have used implicitly the Markov kernel

$$\begin{aligned} K : [0,1]^2 \times \mathcal{B}^1 &\to [0,1] \\ ((a,b), B) &\mapsto A_{a,b}(\mu)(B). \end{aligned} \tag{4.2}$$

Obviously $K((a,b), \cdot\,)$ is a probability measure for fixed $(a,b) \in [0,1]^2$. That $K(\,\cdot\,, B)$ is measurable for $B \in \mathcal{B}^1$ can be proved directly if B is an interval and extended to $\mathcal{B}^1$ using a Dynkin argument. In this notation the distribution of the $W_i^{(n)}$ used for the definition of $[A_{n+1}, B_{n+1}]$ has the following mixed structure (denoting with π_2 the projection onto the second component):

$$\pi_2(K \otimes \mathbb{P}^{(A_n, B_n)}). \tag{4.3}$$

In the case where μ is the uniform distribution on $[0,1]$ leads to random variables "uniformly distributed between two random variables", which were analyzed on its own by van Assche (1987). We have the following distributional relation

$$\mathbb{P}^{(A_{n+1},B_{n+1})\,|\,(A_n,B_n)=(a,b)} = A_{a,b}(\mu) \otimes A_{a,b}(\mu). \tag{4.4}$$

We can also give a pointwise recursion for (A_n, B_n). Let $\{U^{(n)}, V^{(n)} : n \in \mathbb{N}_0\}$ denote an independent family of r.v. with distribution μ. Then

$$\begin{aligned}
\begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} &= \begin{pmatrix} 1-U^{(n)} & U^{(n)} \\ 1-V^{(n)} & V^{(n)} \end{pmatrix} \begin{pmatrix} A_n \\ B_n \end{pmatrix} \\
&= T_n \cdot \ldots \cdot T_0 \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\end{aligned} \tag{4.5}$$

with independent, random stochastic matrices

$$T_i := \begin{pmatrix} 1-U^{(i)} & U^{(i)} \\ 1-V^{(i)} & V^{(i)} \end{pmatrix} \quad \text{for} \quad i \geq 0. \tag{4.6}$$

That the intervals $([A_n, B_n])$ shrink to a point Y a.s. can be shown by standard arguments: It is

$$\begin{aligned}
&\mathbb{E}\,|B_{n+1} - A_{n+1}| \\
&= \int\int |x-y|\; A_{a,b}(\mu) \otimes A_{a,b}(\mu)(d(x,y))\; \mathbb{P}^{(A_n,B_n)}(d(a,b)) \\
&= \int\int |A_{a,b}(x) - A_{a,b}(y)|\; \mu \otimes \mu(d(x,y))\; \mathbb{P}^{(A_n,B_n)}(d(a,b)) \\
&= \int\int |b-a|\,|x-y|\; \mu \otimes \mu(d(x,y))\; \mathbb{P}^{(A_n,B_n)}(d(a,b)) \\
&= \mathbb{E}\,|B_n - A_n| \int |x-y|\; \mu \otimes \mu(d(x,y)) \qquad (4.7) \\
&= \xi\,\mathbb{E}\,|B_n - A_n|.
\end{aligned}$$

The integrand in (4.7) is ≤ 1 and < 1 with probability at least $1/2$. Thus $\xi < 1$. By induction we deduce

$$\mathbb{E}\,|B_n - A_n| = \xi^n. \tag{4.8}$$

Using Markov's inequality it follows

$$\sum_{n=1}^{\infty} \mathbb{P}(|B_n - A_n| > \varepsilon) \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}\,|B_n - A_n|}{\varepsilon} < \infty. \tag{4.9}$$

The Borel-Cantelli Lemma implies $\mathbb{P}(\limsup_{n\to\infty}\{|B_n - A_n| > \varepsilon\}) = 0$. This holds for every $\varepsilon > 0$. Thus $|B_n - A_n| \to 0$ almost surely. Since the segments $[A_n, B_n]$ are nested it follows $\lim_{n\to\infty} B_n = \lim_{n\to\infty} A_n =: Y$ a.s.

Now, we investigate a rate of convergence for the splitting scheme of the following type: Choose $W^{(n)}$ on $[A_n, B_n]$ according to μ. Then

$$W^{(n)} \stackrel{\mathcal{D}}{=} \pi_2(K \otimes \mathbb{P}^{(A_n,B_n)}) \tag{4.10}$$

and $\min\{A_n, B_n\} \leq W^{(n)} \leq \max\{A_n, B_n\}$. The almost sure convergence of A_n, B_n implies $W^{(n)} \to Y$ a.s., in particular $W^{(n)} \to Y$ in distribution. Therefore $\mathcal{L}(W^{(n)})$, which is the *unconditioned* distribution of the points chosen in the nth step to define (A_{n+1}, B_{n+1}), stabilizes to the limit point of the interval splitting scheme. This is also reasonable if we think of choosing the "next" points in the infinityth step: We have to choose between $\lim_{n\to\infty} A_n = Y$ and $\lim_{n\to\infty} B_n = Y$, i.e. we choose Y. From this point of view it is natural to measure the rate of convergence of the splitting scheme in terms of a distance between $W^{(n)}$ and Y.

We start with a distributional recursion for $W^{(n)}$. Observe that the vectors (A_n, B_n) satisfy a two-dimensional recursion. *A priori* it is not obvious that the one-dimensional mixture $W^{(n)}$ contains a sufficient amount of information about the state of the splitting scheme so that a distributional relation only in terms of $W^{(n)}$ can be established.

Lemma 4.2.1 *$(W^{(n)})$ given by (4.10) satisfies the recursion*

$$W^{(n+1)} \stackrel{\mathcal{D}}{=} U + (V-U)W^{(n)} \quad \textit{for all} \quad n \geq 0, \tag{4.11}$$

where $U, V, W^{(n)}$ are independent and $U, V \sim \mu$.

Proof: We expand the distributions of both sides of (4.11). Let $B \in \mathcal{B}^1$.

`Right side of (4.11):` It is

$$\begin{aligned}
&\mathbb{P}(U + (V-U)W^{(n)} \in B)\\
&= \int\int \mathbf{1}_B\Big(u + (v-u)w\Big)\,\mathbb{P}^{W^{(n)}}(dw)\,\mathbb{P}^{(U,V)}(d(u,v))\\
&= \int\int\int \mathbf{1}_B\Big(u + (v-u)w\Big)\,A_{a,b}(\mu)(dw)\\
&\qquad\qquad \mathbb{P}^{T_{n-1}\cdot\ldots\cdot T_0(0,1)^t}(d(a,b))\,\mu\otimes\mu(d(u,v))\\
&= \int\int\int \mathbf{1}_B\Big(u + (v-u)(a + (b-a)w)\Big)\,\mu(dw)\\
&\qquad\qquad \mathbb{P}^{T_{n-1}\cdot\ldots\cdot T_0(0,1)^t}(d(a,b))\,\mu\otimes\mu(d(u,v)).
\end{aligned} \tag{4.12}$$

`Left side of (4.11):` It is

$$\begin{aligned}
&\mathbb{P}(W^{(n+1)} \in B) \\
&= \int\int \mathbf{1}_B(w)\, A_{a,b}(\mu)(dw)\, \mathbb{P}^{T_n\cdot\ldots\cdot T_0(0,1)^t}(d(a,b)) \\
&= \int\int\int \mathbf{1}_B(w)\, A_{a,b}(\mu)(dw)\, \mathbb{P}^{T_n\cdot\ldots\cdot T_1(u,v)^t}(d(a,b))\, \mathbb{P}^{(U,V)}(d(u,v)) \\
&= \int\int\int \mathbf{1}_B(w)\, A_{a,b}(\mu)(dw) \\
&\qquad\qquad \mathbb{P}^{T_n\cdot\ldots\cdot T_1 A'_{u,v}(0,1)^t}(d(a,b))\, \mu\otimes\mu(d(u,v)) \qquad (4.13)
\end{aligned}$$

with

$$\begin{aligned}
A'_{u,v} : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\
\begin{pmatrix} a \\ b \end{pmatrix} &\mapsto (v-u)\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} u \\ u \end{pmatrix}. \qquad (4.14)
\end{aligned}$$

Since the T_i are stochastic matrices we have

$$T_n\cdot\ldots\cdot T_1 \begin{pmatrix} u \\ u \end{pmatrix} = \begin{pmatrix} u \\ u \end{pmatrix}. \qquad (4.15)$$

This implies

$$T_n\cdot\ldots\cdot T_1 A'_{u,v} = A'_{u,v} T_n\cdot\ldots\cdot T_1. \qquad (4.16)$$

Together with (4.13) this yields

$$\begin{aligned}
&\mathbb{P}(W^{(n+1)} \in B) \\
&= \int\int\int \mathbf{1}_B(w)\, A_{a,b}(\mu)(dw) \\
&\qquad\qquad \mathbb{P}^{A'_{u,v} T_n\cdot\ldots\cdot T_1(0,1)^t}(d(a,b))\, \mu\otimes\mu(d(u,v)) \\
&= \int\int\int \mathbf{1}_B(w)\, A_{A'_{u,v}(a,b)}(\mu)(dw) \\
&\qquad\qquad \mathbb{P}^{T_n\cdot\ldots\cdot T_1(0,1)^t}(d(a,b))\, \mu\otimes\mu(d(u,v)) \\
&= \int\int\int \mathbf{1}_B(w)\, A_{(v-u)a+u,(v-u)b+u}(\mu)(dw) \\
&\qquad\qquad \mathbb{P}^{T_n\cdot\ldots\cdot T_1(0,1)^t}(d(a,b))\, \mu\otimes\mu(d(u,v))
\end{aligned}$$

$$= \int\int\int \mathbf{1}_B\Big((v-u)a + u - (v-u)(b-a)w\Big)\, \mu(dw)$$

$$\mathbb{P}^{T_n\cdot\ldots\cdot T_1(0,1)^t}(d(a,b))\ \mu\otimes\mu(d(u,v)) \qquad (4.17)$$

The integrands in (4.17) and (4.12) are identical and $T_n \cdot \ldots \cdot T_1 \sim T_{n-1} \cdot \ldots \cdot T_0$. This implies the assertion. ■

Now, we use equation (4.11) to define the corresponding operator

$$\begin{aligned} S : M^1(\mathbb{R}^1, \mathcal{B}^1) &\to M^1(\mathbb{R}^1, \mathcal{B}^1) \\ \nu &\mapsto \mathcal{L}(U + (V-U)Z), \end{aligned} \qquad (4.18)$$

where U, V, Z are independent, $Z \sim \nu$ and $U, V \sim \mu$.

Lemma 4.2.2 $S : M_{1/2,p} \to M_{1/2,p}$ *is a contraction w.r.t.* ℓ_p *for all* $p \geq 1$*:*

$$\ell_p(S(\lambda), S(\nu)) \leq \|U - V\|_p\, \ell_p(\lambda, \nu) \quad \textit{for all} \quad \lambda, \nu \in M_{1/2,p}, \qquad (4.19)$$

where U, V *are independent with* $U, V \sim \mu$*.*

Proof: Since U, V have pth moments and by independence $\mathbb{E}\,\|S(\nu)\|^p < \infty$ for $\nu \in M_{1/2,p}$ and $p \geq 1$. Furthermore

$$\mathbb{E}\, S(\nu) = \mathbb{E}\, U - \mathbb{E}\,(V-U)\,\mathbb{E}\, Z = 1/2, \qquad (4.20)$$

since $\mathbb{E}\, U = \mathbb{E}\,\mu = 1/2$. So, $S : M_{1/2,p} \to M_{1/2,p}$ is well defined. Let $\lambda, \nu \in M_{1/2,p}$ and Z_1, Z_2 be optimal ℓ_p-couplings of λ, ν. Then with $U, V \sim \mu$, so that $U, V, (Z_1, Z_2)$ are independent

$$\begin{aligned} \ell_p^p(S(\lambda), S(\nu)) &\leq \mathbb{E}\,[(U-V)^p (Z_1 - Z_2)^p] \\ &= \|U - V\|_p^p\, \ell_p^p(\lambda, \nu). \end{aligned} \qquad (4.21)$$

It is $U - V \leq 1$ and $U - V < 1$ with probability at least 1/2. This implies $\|U - V\|_p < 1$. ■

By Banach's fixed point theorem S has a unique fixed point in $M_{1/2,p}$ and $S^n(\nu)$ converges in ℓ_p to this fixed point for every $\nu \in M_{1/2,p}$. Choose $\nu := \mu \in M_{1/2,p}$. It is $W^{(0)} \sim \mu$ and by iterated application of Lemma 4.2.1

$$W^{(n)} = S^n(\mu). \qquad (4.22)$$

Since $W^{(n)} \to Y$ in distribution, where Y is the point to which the interval splitting scheme shrinks we deduce that Y is the fixed point of S,

$$S(Y) \stackrel{\mathcal{D}}{=} Y. \qquad (4.23)$$

Then using (4.19)

$$\begin{aligned}\ell_p(W^{(n)}, Y) &= \ell_p(S(W^{(n-1)}), S(Y)) \\ &\leq \|U - V\|_p \, \ell_p(W^{(n-1)}, Y) \\ &\leq \|U - V\|_p^n \, \ell_p(\mu, Y), \end{aligned} \tag{4.24}$$

which gives the convergence rate of the splitting scheme:

Theorem 4.2.3 *Let μ be a probability measure on $[0,1]$ with $\mathbb{E}\,\mu = 1/2$, $W^{(n)}$ defined by (4.10), and Y the fixed point in $M_{1/2,p}$ of S given by (4.18). Then*

$$\ell_p(W^{(n)}, Y) \leq \ell_p(\mu, Y)\, \|U - V\|_p^n \quad \text{for all} \quad n \geq 0,\ p \geq 1, \tag{4.25}$$

where U, V are independent with $U, V \sim \mu$.

For the special case of $\mu = U_{[0,1]}$, $U_{[0,1]}$ denoting the uniform distribution on the unit interval Y is known to be beta$(2,2)$ distributed. Then (4.25) can be stated more explicitly:

Corollary 4.2.4 *For $\mu = U_{[0,1]}$ in Theorem 4.2.3 it holds*

$$\ell_p(W^{(n)}, Y) \leq \ell_p\Big(U_{[0,1]}, \mathrm{beta}(2,2)\Big) \left(\frac{2}{(p+1)(p+2)}\right)^{n/p} \tag{4.26}$$

for all $n \geq 0$ and $p \geq 1$.

A result of van Assche (1986) on products of random stochastic matrices as arising in (4.5) implies that in the case $\mu = \mathrm{beta}(\alpha, \alpha)$ for some $\alpha > 0$ the limit point Y of the splitting scheme is beta$(2\alpha, 2\alpha)$ distributed.

Vice versa Theorem 4.2.3 applies to products of random stochastic matrices: Consider the matrices

$$T_i := \begin{pmatrix} 1 - U^{(i)} & U^{(i)} \\ 1 - V^{(i)} & V^{(i)} \end{pmatrix} \quad \text{for} \quad i \geq 0, \tag{4.27}$$

where $(U^{(i)}, V^{(i)})$ are independent with some common joint distribution ν. Then denote the product which again is a stochastic matrix by

$$\begin{pmatrix} 1 - A_{n+1} & A_{n+1} \\ 1 - B_{n+1} & B_{n+1} \end{pmatrix} := T_n \cdot \ldots \cdot T_0. \tag{4.28}$$

It is well known that this product converges in distribution if and only if $\mathcal{L}(U^{(i)}, V^{(i)})$ is not concentrated on $\{(1,0),(0,1)\}$; see Rosenblatt (1964) and Sun (1975). The weak limit is of the form

$$\begin{pmatrix} 1 - Y & Y \\ 1 - Y & Y \end{pmatrix}. \tag{4.29}$$

The distribution function G of Y is characterized as the unique solution of the integral equation

$$G(t) = \int_{x>y} G\left(\frac{t-y}{x-y}\right) d\nu(x,y) + \int_{x<y} \left(1 - G_-\left(\frac{t-y}{x-y}\right)\right) d\nu(x,y). \quad (4.30)$$

Here G_- denotes the leftsided limit of G. Now, assume the special case where $U^{(i)}, V^{(i)}$ are independent, identically distributed with distribution μ, and $\mathbb{E}\,\mu = 1/2$. The A_{n+1}, B_{n+1} in (4.28) coincide with the end points A_{n+1}, B_{n+1} of the segments given by (4.5). Since $A_{n+1}, B_{n+1} \sim W^{(n)}$ with $W^{(n)}$ given by (4.10) the rates of convergence of Theorem 4.2.3 can be applied to the products $T_n \cdot \ldots \cdot T_0$. The characterisation of the limit Y as the fixed point of the operator S given in (4.18) in $M_{1/2,p}$ leads to the integral equation (4.30) for this case.

Corollary 4.2.5 *Let (A_n, B_n) be given by the product of the matrices T_i as in (4.27), (4.28) with $\mathcal{L}(U^{(i)}, V^{(i)}) = \mu \otimes \mu$ and $\mathbb{E}\,\mu = 1/2$. Let Y be the fixed point of S given in (4.18) in $M_{1/2,p}$. Then*

$$\ell_p(A_n, Y) = \ell_p(B_n, Y) \le \ell_p(\mu, Y)\, \|U - V\|_p^n \quad (4.31)$$

for all $n \ge 0$, $p \ge 1$, where U, V are independent with $U, V \sim \mu$.

Chapter 5

Affine recursions

All the limiting distributions of the problems discussed so far, i.e. the limit of the normalized cost of a partial match query in quadtrees or K-d trees, the limit of the scaled path length of a random split tree, the limits of the normalized running time of multiple Quickselect or Find and the distribution of the point to which an interval splitting scheme shrinks, occur as the distributional fixed point of a random affine operator of the form

$$X \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A_r X^{(r)} + b. \tag{5.1}$$

Here $X^{(r)} \sim X$ for $r = 1, \dots, K$ and $(A_1, \dots, A_K, b), X^{(1)}, \dots, X^{(K)}$ are independent. In particular the A_i are not necessarily independent and for the problems treated in the previous chapters b can be written as a function of $A_1, \dots, A_K$ (cf. (1.148), (2.22), (3.19) and (4.18)). In this chapter sequences of distributions which occur by an iteration of an analogous multidimensional operator applied to some initial distribution are discussed.

Let L^0 denote an r.v. in $\mathbb{R}^d$. Let $A^{(1)}, \dots, A^{(K)}$ be random $d \times d$ matrices and b a random translation in $\mathbb{R}^d$. Then for $n \geq 1$ define recursively

$$L^n :\stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A^{(r)} L^{n-1,(r)} + b, \tag{5.2}$$

where $(A^{(1)}, \dots, A^{(K)}, b), L^{n-1,(1)}, \dots, L^{n-1,(K)}$ are independent and $L^{n-1,(r)} \sim L^{n-1}$ for $r = 1, \dots, K$.

The special case $K = 1$ in (5.2) of an iteration of a random affine map has been studied intensively in the literature with many respects. Key references are Kesten (1973), Brand (1986), Bougerol and Picard (1992), and Burton and Rösler (1995). The case $K \geq 2$ leads to branching type recursive sequences. In the one dimensional case without the immigration term b and the $A^{(r)}$ being independent and nonnegative this recursion was studied by Mandelbrot (1974) for the analysis of a model of turbulence of Yaglom and Kolmogorov. To this case further contributions on nontrivial fixed points of a corresponding operator, the existence of moments of these fixed points and convergence of (L_n) to the fixed points were made in Kahane and Peyrière (1976) and Guivarc'h (1990). The case $b = 0$, $A^{(r)} \geq 0$ for $r = 1, \dots K$ with dependencies was considered in Holley and Liggett (1981) and Durrett and Liggett (1983) for the purpose of analyzing a problem in infinite particle systems. The case $b = 0$ with deterministic coefficients (and $K = \infty$) was discussed in Rösler (1998). See this paper also for references and an overview on the one-dimensional fixed point equations without immigration term. The general form of the recursion (5.2) in dimension one was treated in Cramer and Rüschendorf (1996). A two-dimensional version of (5.2) with $K = 2$ and $b = 0$ has been considered in Cramer and Rüschendorf (1998).

Most of the investigations mentioned for the one-dimensional case considered problems of stabilization of the sequence (L_n) itself. In this chapter we are concerned with the convergence of scaled versions of (L_n). In the first two sections we consider the L_2-case and use contraction techniques as in Burton and Rösler (1995) and Cramer and Rüschendorf (1998). In the last section we give an approach for the case $K = 1$ based on the concept of Lyapunov exponents.

5.1 A limit theorem for the $\mathbf{L_2}$-case

Consider the distributional recursion (5.2)

$$L^n :\stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A^{(r)} L^{n-1,(r)} + b, \tag{5.3}$$

where $(A^{(1)}, \dots, A^{(K)}, b), L^{n-1,(1)}, \dots, L^{n-1,(K)}$ are independent, L^0 is square integrable, $A^{(1)}, \dots, A^{(r)}$ are random matrices with $\mathbb{E}\, \|A^{(r)}\|^2 < \infty$ for $r = 1, \dots, K$ and b is a random vector in $\mathbb{R}^d$ with $\mathbb{E}\, \|b\|^2 < \infty$. Let

$$C := \sum_{r=1}^{K} A^{(r)}. \tag{5.4}$$

Then the expectation vector of L^n is given by

$$\begin{aligned} \mathbb{E}\, L^n &= \mathbb{E}\, C\, \mathbb{E}\, L^{n-1} + \mathbb{E}\, b \\ &= (\mathbb{E}\, C)^n\, \mathbb{E}\, L^0 + \sum_{k=0}^{n-1} (\mathbb{E}\, C)^k\, \mathbb{E}\, b. \end{aligned} \tag{5.5}$$

We may assume $\mathrm{Var}(L_i^n) > 0$ for all $1 \le i \le d$ and $n \in \mathbb{N}$. This condition is easy to check and unsatisfied only in very special cases. For the normalization of the process (L^n) define

$$V_n := \mathrm{diag}(\mathrm{Var}(L_1^n)^{-1/2}, \ldots, \mathrm{Var}(L_d^n)^{-1/2}) \quad \text{for} \quad n \in \mathbb{N} \tag{5.6}$$

where diag denotes the diagonal matrix with the given numbers on the diagonal. Consider the scaled quantities

$$\widetilde{L}^n := V_n(L^n - \mathbb{E}\, L^n) \quad \text{and} \tag{5.7}$$

$$A^{(r),\{n\}} := V_n A^{(r)} V_{n-1}^{-1} \quad \text{for} \quad n \ge 1. \tag{5.8}$$

So we normalize L^n by the components not changing its covariance structure:

$$\widetilde{L}_i^n = \frac{L_i^n - \mathbb{E}\, L_i^n}{\mathrm{Var}(L_i^n)^{1/2}} \quad \text{for} \quad 1 \le i \le d, \tag{5.9}$$

$$A_{ij}^{(r),\{n\}} = \left(\frac{\mathrm{Var}(L_j^{n-1})}{\mathrm{Var}(L_i^n)} \right)^{1/2} A_{ij}^{(r)} \quad \text{for} \quad 1 \le i,j \le d. \tag{5.10}$$

This leads to the following modified recursion for $(\widetilde{L}^n)$:

Lemma 5.1.1 *In the given notation the scaled sequence $(\widetilde{L}^n)$ satisfies the modified recursion*

$$\widetilde{L}^n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A^{(r),\{n\}} \widetilde{L}^{n-1,(r)} + V_n(b - \mathbb{E}\, b) + V_n(C - \mathbb{E}\, C)\, \mathbb{E}\, L^{n-1}, \tag{5.11}$$

where $\widetilde{L}^{n-1,(1)}, \ldots, \widetilde{L}^{n-1,(K)}$ are independent copies of $\widetilde{L}^{n-1}$ and independent of $(A^{(1)}, \ldots, A^{(K)}, b)$.

Proof: It is

$$\begin{aligned}
&\sum_{r=1}^{K} A^{(r),\{n\}} \widetilde{L}^{n-1,(r)} \\
&= \sum_{r=1}^{K} V_n A^{(r)} V_{n-1}^{-1} V_{n-1} (L^{n-1,(r)} - \mathbb{E}\, L^{n-1,(r)}) \\
&= V_n \left[\sum_{r=1}^{K} A^{(r)} L^{n-1,(r)} - \left(\sum_{r=1}^{K} A^{(r)} \right) \mathbb{E}\, L^{n-1} \right] \\
&\stackrel{\mathcal{D}}{=} V_n [L^n - b - C\, \mathbb{E}\, L^{n-1}]
\end{aligned} \tag{5.12}$$

$$
\begin{aligned}
&= V_n[L^n - \mathbb{E}\, L^n + \mathbb{E}\, L^n - b - C\,\mathbb{E}\, L^{n-1}] \\
&= V_n(L^n - \mathbb{E}\, L^n) + V_n(\mathbb{E}\, C\,\mathbb{E}\, L^{n-1} + \mathbb{E}\, b - b - C\,\mathbb{E}\, L^{n-1}) \qquad (5.13) \\
&= \widetilde{L}^n - V_n(b - \mathbb{E}\, b) - V_n(C - \mathbb{E}\, C)\,\mathbb{E}\, L^{n-1}.
\end{aligned}
$$

For (5.12) the original recursion (5.3) and for (5.13) the relation (5.5) have been used. ■

In order to define a limiting form of the modified recursion (5.11) assume the existence of the following limits (here $\lim a_n = \infty$ denotes that the sequence (a_n) is definite divergence)

$$\lim_{n\to\infty} \operatorname{Var}(L_i^n) \in (0, \infty], \tag{5.14}$$

$$\lim_{n\to\infty} \left(\frac{\operatorname{Var}(L_j^{n-1})}{\operatorname{Var}(L_i^n)}\right)^{1/2} =: c_{ij}, \tag{5.15}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}\, L_i^n}{(\operatorname{Var}(L_i^n))^{1/2}} =: \gamma_i. \tag{5.16}$$

These assumptions enable us to define limiting objects corresponding to $A^{(r),\{n\}}$ and V_n:

$$A^{(r),\infty} := \lim_{n\to\infty} V_n A^{(r)} V_{n-1}^{-1} = \lim_{n\to\infty} A^{(r),\{n\}}, \quad r = 1, \ldots, K, \tag{5.17}$$

$$C^\infty := \sum_{r=1}^{K} A^{(r),\infty}, \tag{5.18}$$

$$V_\infty := \operatorname{diag}(\vartheta_1, \ldots, \vartheta_d) := \lim_{n\to\infty} V_n. \tag{5.19}$$

Now the limiting operator of the recursion (5.11) is given by

$$
\begin{aligned}
&T : M^1(\mathbb{R}^d, \mathcal{B}^d) \to M^1(\mathbb{R}^d, \mathcal{B}^d) \qquad (5.20) \\
&\mu \mapsto \mathcal{L}\left(\sum_{r=1}^{K} A^{(r),\infty} Z^{(r)} + V_\infty(b - \mathbb{E}\, b) + (C^\infty - \mathbb{E}\, C^\infty)\gamma\right)
\end{aligned}
$$

where $(A^{(1)}, \ldots, A^{(K)}, b), Z^{(1)}, \ldots, Z^{(K)}$ are independent and $Z^{(r)} \sim \mu$ for $r = 1, \ldots, K$. Observe that the last summand in (5.11) is $V_n(C - \mathbb{E}\, C)\,\mathbb{E}\, L^{n-1} = V_n(C - \mathbb{E}\, C)V_{n-1}^{-1}V_{n-1}\,\mathbb{E}\, L^{n-1}$. T satisfies the following Lipschitz condition:

Lemma 5.1.2 *$T : (M^d_{0,2}, \ell_2) \to (M^d_{0,2}, \ell_2)$ with T given by (5.20) is a Lipschitz map:*

$$\ell_2(T(\mu), T(\nu)) \le \xi\, \ell_2(\mu,\nu) \quad \text{for all} \quad \mu,\nu \in M^d_{0,2},$$

$$\xi = \left\| \sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty} \right] \right\|_{op}^{1/2}. \tag{5.21}$$

Proof: Obviously by the valid independencies and L_2-assumptions on the coefficients it holds $\mathbb{E}\,\|T(\mu)\|^2 < \infty$. It is

$$\begin{aligned} &\mathbb{E}\left[\sum_{r=1}^{K} A^{(r),\infty} Z^{(r)} + V_\infty (b - \mathbb{E}\, b) + (C^\infty - \mathbb{E}\, C^\infty)\gamma \right] \\ &= \sum_{r=1}^{K} \mathbb{E}\, A^{(r),\infty}\, \mathbb{E}\, Z^{(r)} + V_\infty\, \mathbb{E}\,(b - \mathbb{E}\, b) + \mathbb{E}\,(C^\infty - \mathbb{E}\, C^\infty)\gamma \\ &= 0 \end{aligned} \tag{5.22}$$

since the $Z^{(r)}$ are centered, so $T : M^d_{0,2} \to M^d_{0,2}$ is well defined. Let $\mu,\nu \in M^d_{0,2}$ and (U,V), $(U^{(1)}, V^{(1)}), \ldots, (U^{(K)}, V^{(K)})$ be optimal ℓ_2-couplings of μ,ν so that $(A^{(1)}, \ldots, A^{(K)}, b)$, (U,V), $(U^{(1)}, V^{(1)}), \ldots, (U^{(K)}, V^{(K)})$ are independent. Then

$$\begin{aligned} &\ell_2^2(T(\mu), T(\nu)) \\ &\le \mathbb{E} \left\| \sum_{r=1}^{K} A^{(r),\infty}(U^{(r)} - V^{(r)}) \right\|^2 \\ &= \mathbb{E} \sum_{r=1}^{K} \|A^{(r),\infty}(U^{(r)} - V^{(r)})\|^2 \\ &\quad + \sum_{\substack{q,r=1 \\ q \neq r}}^{K} \mathbb{E} \left\langle A^{(q),\infty}(U^{(q)} - V^{(q)}), A^{(r),\infty}(U^{(r)} - V^{(r)}) \right\rangle \qquad (5.23) \\ &= \mathbb{E} \sum_{r=1}^{K} \left\langle U^{(r)} - V^{(r)}, \left(A^{(r),\infty}\right)^t A^{(r),\infty}(U^{(r)} - V^{(r)}) \right\rangle \\ &= \sum_{r=1}^{K} \mathbb{E} \left\langle U^{(r)} - V^{(r)}, \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty} \right] (U^{(r)} - V^{(r)}) \right\rangle \qquad (5.24) \\ &= \mathbb{E} \left\langle U - V, \left(\sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty} \right] \right)(U - V) \right\rangle \end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}\left[\|U-V\| \cdot \left\|\left(\sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty}\right]\right)(U-V)\right\|\right] &(5.25)\\
&\leq \mathbb{E}\left[\left\|\sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty}\right]\right\|_{op} \|U-V\|^2\right] &(5.26)\\
&= \left\|\sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty}\right]\right\|_{op} \ell_2^2(\mu,\nu).
\end{aligned}$$

The second summand in (5.23) is zero by independence and $\mathbb{E}\,(U^{(r)} - V^{(r)}) = 0$, the additional expectation in (5.24) is justified by independence. ■

Observe that the estimates (5.25) and (5.26) are sharp simultaneously:

$$S := \sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty}\right] \tag{5.27}$$

is a symmetric, nonnegative definite matrix. Therefore S is diagonalizable with nonnegative eigenvalues. Let $\lambda \geq 0$ denote the greatest of the eigenvalues of S. Then $\|S\|_{op} = \lambda$. Let u be a corresponding eigenvector and define $\varrho := (1/2)\delta_{\{-u\}} + (1/2)\delta_{\{u\}}$. Then equality in (5.25) and (5.26) holds for the measures $\varrho, \delta_{\{0\}} \in M_{0,2}^d$: Let U be an r.v. with $U \sim \varrho$ and $V = 0$. Then U, V are optimal ℓ_2-couplings of ϱ and $\delta_{\{0\}}$ since V is deterministic. Furthermore

$$\begin{aligned}
\mathbb{E}\left\langle U-V, S(U-V)\right\rangle &= \mathbb{E}\left\langle U, SU\right\rangle = \mathbb{E}\left\langle U, \lambda U\right\rangle &(5.28)\\
&= \lambda\,\mathbb{E}\,\|U\|^2 = \lambda\,\mathbb{E}\,\|U-V\|^2\\
&= \|S\|_{op}\,\ell_2^2(\varrho, \delta_{\{0\}}).
\end{aligned}$$

(5.28) is valid since U is a.s. an eigenvector of S to the eigenvalue λ. Therefore (5.25) and (5.26) are sharp simultaneously.

Theorem 5.1.3 (Limit Theorem) *Let (L^n) be the sequence defined by (5.3) where the initial distribution, random matrices and the immigration term are square integrable. Assume the conditions (5.14)–(5.16) and*

$$\xi = \left\|\sum_{r=1}^{K} \mathbb{E}\left[\left(A^{(r),\infty}\right)^t A^{(r),\infty}\right]\right\|_{op}^{1/2} < 1. \tag{5.29}$$

Then the scaled version $(\widetilde{L}^n)$ given by (5.7) converges to the unique fixed point $\widetilde{L}$ in $M_{0,2}^d$ of the limiting operator T given by (5.20):

$$\ell_2(\widetilde{L}^n, \widetilde{L}) \to 0. \tag{5.30}$$

Proof: First, we proof a reduction estimate of the form

$$\ell_2^2(\widetilde{L}^n, \widetilde{L}) \leq \xi^2\, \ell_2^2(\widetilde{L}^{n-1}, \widetilde{L}) + b_n, \quad \text{for all} \quad n \in \mathbb{N}, \tag{5.31}$$

with a sequence (b_n) converging to zero. Let $\widetilde{L}^{n-1,(1)}, \dots, \widetilde{L}^{n-1,(K)}$ be distributional copies of $\widetilde{L}^{n-1}$, $Z^{(1)}, \dots, Z^{(K)} \sim \mu$ and $(\widetilde{L}^{n-1,(r)}, Z^{(r)})$ are optimal ℓ_2-couplings for $r = 1, \dots, K$. Furthermore let $(A^{(1)}, \dots, A^{(K)}, b)$, $(\widetilde{L}^{n-1,(1)}, Z^{(1)}), \dots, (\widetilde{L}^{n-1,(K)}, Z^{(K)})$ be independent. Denote the centered versions of b, C and C^∞ by $b^c := b - \mathbb{E}\, b$, $C^c := C - \mathbb{E}\, C$ and $C^{c,\infty} := C^\infty - \mathbb{E}\, C^\infty$. Using the modified recursion (5.11) and the fixed point relation we deduce

$$\begin{aligned}
&\ell_2^2(\widetilde{L}^n, \widetilde{L}) \\
&\leq \mathbb{E}\Bigg\| \sum_{r=1}^K A^{(r),\{n\}} \widetilde{L}^{n-1,(r)} + V_n(b - \mathbb{E}\, b) + V_n(C - \mathbb{E}\, C)\, \mathbb{E}\, L^{n-1} \\
&\qquad - \left(\sum_{r=1}^K A^{(r),\infty} Z^{(r)} + V_\infty(b - \mathbb{E}\, b) + (C^\infty - \mathbb{E}\, C^\infty)\gamma \right) \Bigg\|^2 \\
&= \mathbb{E}\Bigg\| \sum_{r=1}^K A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}) + \sum_{r=1}^K (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \\
&\qquad + (V_n - V_\infty) b^c + V_n C^c\, \mathbb{E}\, L^{n-1} - C^{c,\infty}\gamma \Bigg\|^2 \\
&\leq \mathbb{E}\Bigg\| \sum_{r=1}^K A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}) \Bigg\|^2 + \mathbb{E}\Bigg\| \sum_{r=1}^K (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \Bigg\|^2 \\
&\quad + \mathbb{E}\, \|(V_n - V_\infty) b^c\|^2 + \mathbb{E}\, \|V_n C^c\, \mathbb{E}\, L^{n-1} - C^{c,\infty}\gamma\|^2 \\
&\quad + 2\, \mathbb{E} \left\langle \sum_{r=1}^K A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}), \sum_{r=1}^K (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \right\rangle \\
&\quad + 2\, \mathbb{E} \left\langle (V_n - V_\infty) b^c, V_n C^c\, \mathbb{E}\, L^{n-1} - C^{c,\infty}\gamma \right\rangle.
\end{aligned} \tag{5.32}$$

The other inner products vanish by independence and $\mathbb{E}\, \widetilde{L}^{n-1,(r)} = \mathbb{E}\, Z^{(r)} = 0$. For the last summand in (5.32) we use $\langle a, b \rangle \leq \|a\|^2 + \|b\|^2$. The first summand in (5.32) is bounded from above by $\xi^2\, \ell_2^2(\widetilde{L}^{n-1}, \widetilde{L})$, which can be obtained by a calculation similar to the proof of Lemma 5.1.2. This yields

$$\ell_2^2(\widetilde{L}^n, \widetilde{L})$$
$$\leq \xi^2 \ell_2^2(\widetilde{L}^{n-1}, \widetilde{L})$$
$$+ \mathbb{E}\left\| \sum_{r=1}^{K} (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \right\|^2 \tag{5.33}$$
$$+ 3\,\mathbb{E}\, \|(V_n - V_\infty) b^c\|^2 \tag{5.34}$$
$$+ 3\,\mathbb{E}\, \|V_n C^c\, \mathbb{E}\, L^{n-1} - C^{c,\infty}\gamma\|^2 \tag{5.35}$$
$$+ 2\,\mathbb{E}\left\langle \sum_{r=1}^{K} A^{(r),\infty} (\widetilde{L}^{n-1,(r)} - Z^{(r)}), \sum_{r=1}^{K} (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \right\rangle. \tag{5.36}$$

It remains to prove that the summands in (5.33)–(5.36) are converging to zero. First observe that

$$\mathbb{E}\, \|A^{(r),\{n\}} - A^{(r),\infty}\|^2 \tag{5.37}$$
$$= \mathbb{E} \sum_{i,j=1}^{d} (A_{ij}^{(r),\{n\}} - A_{ij}^{(r),\infty})^2$$
$$= \sum_{i,j=1}^{d} \mathbb{E}\,[A_{ij}^2] \left(\left(\frac{\mathrm{Var}(L_j^{n-1})}{\mathrm{Var}(L_i^n)} \right)^{1/2} - c_{ij} \right)^2 \longrightarrow 0$$

by (5.15), in particular

$$\mathbb{E}\, \|A^{(r),\{n\}} - A^{(r),\infty}\|^2 = \mathbb{E}\, \|A^{(r)}\|^2 \cdot o(1) \quad \text{for} \quad r = 1, \ldots, K. \tag{5.38}$$

`Summand (5.33):`

$$\mathbb{E}\left\| \sum_{r=1}^{K} (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \right\|^2$$
$$= \mathbb{E} \sum_{r=1}^{K} \left\| (A^{(r),\{n\}} - A^{(r),\infty}) \widetilde{L}^{n-1,(r)} \right\|^2$$
$$\leq \sum_{r=1}^{K} \mathbb{E} \left[\|A^{(r),\{n\}} - A^{(r),\infty}\|^2 \|\widetilde{L}^{n-1,(r)}\|^2 \right]$$
$$= \sum_{r=1}^{K} \mathbb{E}\, \|A^{(r),\{n\}} - A^{(r),\infty}\|^2\, \mathbb{E}\, \|\widetilde{L}^{n-1,(r)}\|^2$$
$$= \sum_{r=1}^{K} \mathbb{E}\, \|A^{(r)}\|^2\, o(1)\, d \longrightarrow 0 \quad \text{for} \quad n \to \infty, \tag{5.39}$$

where $\mathbb{E}\,\|\widetilde{L}^{n-1,(r)}\|^2 = d$ has been used, which is caused by the scaling (5.9).

`Summand (5.34):`

$$\mathbb{E}\,\|(V_n - V_\infty)b^c\|^2 \le \max_{1\le i\le d}\left\{\frac{1}{(\mathrm{Var}(L_i^n))^{1/2}} - \vartheta_i\right\}^2 \mathbb{E}\,\|b^c\|^2 \longrightarrow 0 \qquad (5.40)$$

by (5.14) and (5.19).

`Summand (5.35):`

$$\begin{aligned}
&\mathbb{E}\,\|V_n C^c\,\mathbb{E}\,L^{n-1} - C^{c,\infty}\gamma\|^2\\
&= \mathbb{E}\,\|V_n C^c V_{n-1}^{-1} V_{n-1}\,\mathbb{E}\,L^{n-1} - C^{c,\infty}\gamma\|^2\\
&= \mathbb{E}\,\|(V_n C^c V_{n-1}^{-1} - C^{c,\infty})V_{n-1}\,\mathbb{E}\,L^{n-1} + C^{c,\infty}(V_{n-1}\,\mathbb{E}\,L^{n-1} - \gamma)\|^2\\
&\le 3\Big(\mathbb{E}\,\|(V_n C^c V_{n-1}^{-1} - C^{c,\infty})V_{n-1}\,\mathbb{E}\,L^{n-1}\|^2\\
&\qquad + \mathbb{E}\,\| + C^{c,\infty}(V_{n-1}\,\mathbb{E}\,L^{n-1} - \gamma)\|^2\Big)\\
&\le 3\Big(\mathbb{E}\,\big[\|C^{c,\{n\}} - C^{c,\infty}\|^2 \|V_{n-1}\,\mathbb{E}\,L^{n-1}\|^2\big]\\
&\qquad + \mathbb{E}\,\big[\|C^{c,\infty}\|^2 \|V_{n-1}\,\mathbb{E}\,L^{n-1} - \gamma\|^2\big]\Big)\\
&= 3\Big(o(1)\,\|C^c\|^2 \|V_{n-1}\,\mathbb{E}\,L^{n-1}\|^2 + \mathbb{E}\,\|C^{c,\infty}\|^2 \|V_{n-1}\,\mathbb{E}\,L^{n-1} - \gamma\|^2\Big)\\
&\longrightarrow 0 \quad \text{for} \quad n \to \infty, \qquad (5.41)
\end{aligned}$$

since $\|V_{n-1}\,\mathbb{E}\,L^{n-1}\|^2$ is bounded by (5.16) and $\|V_{n-1}\,\mathbb{E}\,L^{n-1} - \gamma\| \to 0$ by (5.16).

`Summand (5.36):`

$$\begin{aligned}
&\mathbb{E}\left\langle \sum_{r=1}^K A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}), \sum_{q=1}^K (A^{(q),\{n\}} - A^{(q),\infty})\widetilde{L}^{n-1,(q)} \right\rangle\\
&= \mathbb{E}\sum_{q,r=1}^K \left\langle A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}), (A^{(q),\{n\}} - A^{(q),\infty})\widetilde{L}^{n-1,(q)} \right\rangle\\
&= \mathbb{E}\sum_{r=1}^K \left\langle A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)}), (A^{(r),\{n\}} - A^{(r),\infty})\widetilde{L}^{n-1,(r)} \right\rangle\\
&\le \sum_{r=1}^K \mathbb{E}\left[\|A^{(r),\infty}(\widetilde{L}^{n-1,(r)} - Z^{(r)})\| \cdot \|(A^{(r),\{n\}} - A^{(r),\infty})\widetilde{L}^{n-1,(r)}\|\right]
\end{aligned}$$

$$\begin{aligned} &\leq \sum_{r=1}^{K} o(1)\,\mathbb{E}\,[\|A^{(r),\infty}\|\,\|A^{(r)}\|]\Big(\mathbb{E}\,\|\widetilde{L}^{n-1,(r)}\|^2 + \mathbb{E}\,[\|Z^{(r)}\|\,\|\widetilde{L}^{n-1,(r)}\|]\Big) \\ &\longrightarrow 0, \end{aligned} \tag{5.42}$$

since $\mathbb{E}\,\|\widetilde{L}^{n-1,(r)}\|^2 = d$ and using the Cauchy-Schwarz inequality

$$\begin{aligned} \mathbb{E}\,[\|Z^{(r)}\|\,\|\widetilde{L}^{n-1,(r)}\|] &\leq \mathbb{E}\,[\|Z^{(r)}\|^2]^{1/2}\,\mathbb{E}\,[\|\widetilde{L}^{n-1,(r)}\|^2]^{1/2} \\ &\leq \text{const}\ \sqrt{d}. \end{aligned} \tag{5.43}$$

Altogether this implies

$$\ell_2^2(\widetilde{L}^n, \widetilde{L}) \leq \xi^2\,\ell_2^2(\widetilde{L}^{n-1}, \widetilde{L}) + b_n \tag{5.44}$$

for $n \geq 1$. Iteration for this reduction inequality leads to

$$\ell_2^2(\widetilde{L}^n, \widetilde{L}) \leq \xi^{2n}\,\ell_2^2(\widetilde{L}^0, \widetilde{L}) + \sum_{k=0}^{n-1} \xi^{2k} b_{n-k}. \tag{5.45}$$

Since $\xi < 1$ and $b_n \to 0$ we can directly deduce

$$\sum_{k=0}^{n-1} \xi^{2k} b_{n-k} \longrightarrow 0 \quad \text{for} \quad n \to \infty \tag{5.46}$$

(see Cramer (1995, Lemma 6.30)). This implies

$$\ell_2^2(\widetilde{L}^n, \widetilde{L}) \longrightarrow 0 \quad \text{for} \quad n \to \infty. \tag{5.47}$$

■

5.2 Covariance structure

For the application of Theorem 5.1.3 it is necessary to check the conditions (5.14)–(5.16). The asymptotic of the mean vector $\mathbb{E}\,L^n$ is given by (5.5)

$$\mathbb{E}\,L^n = (\,\mathbb{E}\,C)^n\,\mathbb{E}\,L^0 + \sum_{k=0}^{n-1} (\,\mathbb{E}\,C)^k\,\mathbb{E}\,b. \tag{5.48}$$

For the mean, therefore, the computational complexity reduces to the derivation of the powers of the matrix $\mathbb{E}\,C$. For the asymptotics of the variances $\mathrm{Var}(L_1^n), \ldots, \mathrm{Var}(L_d^n)$ we have to investigate the whole covariance matrix $\mathrm{Cov}(L^n)$. We use the following notation: For an $n \times m$ matrix $A = (a_{ij})$ denote by $A_s := (a_{s1}, \ldots, a_{sm})^t$ the sth row vector of A for $1 \leq s \leq n$. If X is an r.v. in $\mathbb{R}^d$, $\mathrm{Cov}(X)$ denotes the matrix $(\mathrm{Cov}(X_i, X_j))_{i,j=1}^d$. For two r.v. X, Y in $\mathbb{R}^d$, $\mathbb{R}^{d'}$ respectively, $\mathrm{Cov}(X, Y)$ denotes the $d \times d'$ matrix $(\mathrm{Cov}(X_i, Y_j))$, $1 \leq i \leq d, 1 \leq j \leq d'$.

Lemma 5.2.1 *Let (L^n) satisfy the recursion (5.3) and $C = \sum_{r=1}^K A^{(r)}$. Then for $1 \le s, t \le d$*

$$\begin{aligned}(\mathrm{Cov}(L^n))_{st} &= \sum_{r=1}^K \mathbb{E}\left\langle \mathrm{Cov}(L^{n-1})A_s^{(r)}, A_t^{(r)}\right\rangle \\ &\quad + \left\langle \mathrm{Cov}(C_s, C_t)\,\mathbb{E}\,L^{n-1}, \mathbb{E}\,L^{n-1}\right\rangle \\ &\quad + \left\langle \mathrm{Cov}(C_s, b_t) + \mathrm{Cov}(C_t, b_s), \mathbb{E}\,L^{n-1}\right\rangle \\ &\quad + (\mathrm{Cov}(b))_{st} \end{aligned} \tag{5.49}$$

Proof: Using linearity we deduce

$$\begin{aligned}(\mathrm{Cov1}(L^n))_{st} &= \mathrm{Cov}\left(\sum_{r=1}^K \left\langle A_s^{(r)}, L^{n-1,(r)}\right\rangle + b_s, \sum_{q=1}^K \left\langle A_t^{(q)}, L^{n-1,(q)}\right\rangle + b_t\right) \\ &= \sum_{q,r=1}^K \mathrm{Cov}\left(\left\langle A_s^{(r)}, L^{n-1,(r)}\right\rangle, \left\langle A_t^{(q)}, L^{n-1,(q)}\right\rangle\right) && (5.50) \\ &\quad + \sum_{r=1}^K \mathrm{Cov}\left(\left\langle A_s^{(r)}, L^{n-1,(r)}\right\rangle + b_t\right) && (5.51) \\ &\quad + \sum_{q=1}^K \mathrm{Cov}\left(b_s, \left\langle A_t^{(q)}, L^{n-1,(q)}\right\rangle\right) && (5.52) \\ &\quad + \mathrm{Cov}\left(b_s, b_t\right). && (5.53)\end{aligned}$$

Let $A, \widetilde{A}, L, \widetilde{L}$ be square integrable r.v. in $\mathbb{R}^1$ so that $(A, \widetilde{A}), (L, \widetilde{L})$ are independent. Then a direct calculation leads to the formula

$$\mathrm{Cov}(AL, \widetilde{A}\widetilde{L}) = \mathrm{Cov}(L, \widetilde{L})\,\mathbb{E}\,A\,\mathbb{E}\,\widetilde{A} + \mathrm{Cov}(A, \widetilde{A})\,\mathbb{E}\,[L\widetilde{L}]. \tag{5.54}$$

With this identity we first expand the summand (5.50)

$$\begin{aligned}&\sum_{q,r=1}^K \mathrm{Cov}\left(\left\langle A_s^{(r)}, L^{n-1,(r)}\right\rangle, \left\langle A_t^{(q)}, L^{n-1,(q)}\right\rangle\right) \\ &= \sum_{q,r=1}^K \sum_{i,j=1}^d \mathrm{Cov}\left(A_{si}^{(r)} L_i^{n-1,(r)}, A_{tj}^{(q)} L_j^{n-1,(q)}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{q,r=1}^{K} \sum_{i,j=1}^{d} \biggl(\mathrm{Cov}\left(L_i^{n-1,(r)}, L_j^{n-1,(q)} \right) \mathbb{E}\left[A_{si}^{(r)} A_{tj}^{(q)} \right] \\
&\qquad\qquad + \mathrm{Cov}\left(A_{si}^{(r)}, A_{tj}^{(q)} \right) \mathbb{E}\, L_i^{n-1,(r)}\, \mathbb{E}\, L_j^{n-1,(q)} \biggr) \\
&= \sum_{q,r=1}^{K} \sum_{i,j=1}^{d} \biggl(\delta_{rq} \mathrm{Cov}\left(L_i^{n-1}, L_j^{n-1} \right) \mathbb{E}\left[A_{si}^{(r)} A_{tj}^{(q)} \right] \\
&\qquad\qquad + \mathrm{Cov}\left(A_{si}^{(r)}, A_{tj}^{(q)} \right) \mathbb{E}\, L_i^{n-1}\, \mathbb{E}\, L_j^{n-1} \biggr) \\
&= \sum_{r=1}^{K} \mathbb{E}\left[\sum_{i,j=1}^{d} \mathrm{Cov}\left(L_i^{n-1}, L_j^{n-1} \right) A_{si}^{(r)} A_{tj}^{(r)} \right] \\
&\quad + \sum_{i,j=1}^{d} \left(\sum_{q,r=1}^{K} \mathrm{Cov}\left(A_{si}^{(r)}, A_{tj}^{(q)} \right) \right) \mathbb{E}\, L_i^{n-1}\, \mathbb{E}\, L_j^{n-1} \\
&= \sum_{r=1}^{K} \mathbb{E}\left[\left\langle \mathrm{Cov}\left(L^{n-1} \right) A_s^{(r)}, A_t^{(r)} \right\rangle \right] \\
&\quad + \sum_{i,j=1}^{d} \mathrm{Cov}\left(C_{si}, C_{tj} \right) \mathbb{E}\, L_i^{n-1}\, \mathbb{E}\, L_j^{n-1} \\
&= \sum_{r=1}^{K} \mathbb{E}\left[\left\langle \mathrm{Cov}\left(L^{n-1} \right) A_s^{(r)}, A_t^{(r)} \right\rangle \right] \\
&\quad + \left\langle \mathrm{Cov}(C_s, C_t)\, \mathbb{E}\, L^{n-1},\, \mathbb{E}\, L^{n-1} \right\rangle. \qquad (5.55)
\end{aligned}
$$

The summand (5.51) can be expanded to

$$
\begin{aligned}
&\sum_{r=1}^{K} \mathrm{Cov}\Bigl(\left\langle A_s^{(r)}, L^{n-1,(r)} \right\rangle + b_t \Bigr) \\
&\quad = \sum_{r=1}^{K} \sum_{i=1}^{d} \mathrm{Cov}\left(A_{si}^{(r)} L_i^{n-1,(r)}, b_t \right) \\
&\quad = \sum_{r=1}^{K} \sum_{i=1}^{d} \mathbb{E}\, L_i^{n-1,(r)} \mathrm{Cov}\left(A_{si}^{(r)}, b_t \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{d} \mathbb{E}\, L_i^{n-1} \sum_{r=1}^{K} \mathrm{Cov}\left(A_{si}^{(r)}, b_t\right) \\
&= \sum_{i=1}^{d} \mathbb{E}\, L_i^{n-1} \mathrm{Cov}\left(C_{si}, b_t\right) \\
&= \left\langle \mathrm{Cov}(C_s, b_t),\, \mathbb{E}\, L^{n-1} \right\rangle. \qquad (5.56)
\end{aligned}
$$

Summand (5.52) has the analogous representation. So, (5.51) and (5.52) together make the contribution

$$\left\langle \mathrm{Cov}(C_s, b_t) + \mathrm{Cov}(C_t, b_s),\, \mathbb{E}\, L^{n-1} \right\rangle \tag{5.57}$$

to the expansion of $(\mathrm{Cov}(L^n))_{st}$. ■

Our aim is to give a formula for $\mathrm{Var}(L_i^n)$ only in terms of the given quantities $A^{(1)}, \ldots, A^{(K)}, b, L^0$. For an iteration of formula (5.49) we introduce the following notation: For a given $d \times d$ matrix $M = (m_{ij})$ denote by $M^V \in \mathbb{R}^{d^2}$ the vector

$$M^V_{(i-1)d+j} := m_{ij}, \quad \text{for all} \quad 1 \le i, j \le d. \tag{5.58}$$

This means that we write the matrix M row by row into the vector M^V. We also use the notation $(\mathrm{Cov}(\,\cdot\,))^V = \mathrm{Cov}^V(\,\cdot\,)$. Define the $d^2 \times d^2$ matrices P,Q and the $d^2 \times d$ matrix R by

$$
\begin{aligned}
P_{(s-1)d+t,(i-1)d+j} &:= \sum_{r=1}^{K} \mathbb{E}\left[A_{si}^{(r)} A_{tj}^{(r)}\right], && (5.59)\\
Q_{(s-1)d+t,(i-1)d+j} &:= \sum_{q,r=1}^{K} \mathrm{Cov}\left(A_{si}^{(r)}, A_{tj}^{(r)}\right), && (5.60)\\
R_{(s-1)d+t,j} &:= \sum_{r=1}^{K} \mathrm{Cov}\left(A_{si}^{(r)}, b_t\right) + \mathrm{Cov}\left(A_{tj}^{(r)}, b_s\right), && (5.61)
\end{aligned}
$$

for $1 \le i, j, s, t \le d$. Furthermore denote by M_n the $d \times d$ matrix

$$M_n := \mathbb{E}\, L^n \cdot (\mathbb{E}\, L^n)^t. \tag{5.62}$$

By (5.48) M_n is expressible only in terms of $A^{(1)}, \ldots, A^{(K)}, b, L^0$. In this notation the recurrence (5.49) reads

$$\mathrm{Cov}^V(L^n) = P\,\mathrm{Cov}^V(L^{n-1}) + QM_{n-1}^V + R\,\mathbb{E}\, L^{n-1} + \mathrm{Cov}^V(b). \tag{5.63}$$

Iteration of this formula leads to an explicit expansion of the covariance matrix $\mathrm{Cov}(L^n)$:

Theorem 5.2.2 *The covariance matrix $Cov(L^n)$ of L^n satisfying the recursion (5.3) in the notions (5.58)–(5.62) has the expansion*

$$\begin{aligned}&\mathrm{Cov}^V(L^n)\\ &= P^n\mathrm{Cov}^V(L^0)+\sum_{k=1}^{n}P^{k-1}\Big(QM_{n-k}^V+R\,\mathbb{E}\,L^{n-k}+\mathrm{Cov}^V(b)\Big).\end{aligned} \tag{5.64}$$

This means that the computational difficulty to determine the asymptotics of $\mathbb{E}\,L^n$ and $\mathrm{Var}(L_1^n),\dots,\mathrm{Var}(L_d^n)$ reduces to calculate the powers of the matrices $\mathbb{E}\,C$ and P given by

$$(\mathbb{E}\,C)_{ij} = \sum_{r=1}^{K}\mathbb{E}\left[A_{ij}^{(r)}\right], \tag{5.65}$$

$$P_{(s-1)d+t,(i-1)d+j} = \sum_{r=1}^{K}\mathbb{E}\left[A_{si}^{(r)}A_{tj}^{(r)}\right] \tag{5.66}$$

for $1 \le i,j,s,t \le d$. A concrete example in the two-dimensional case for the derivation of these asymptotics leading to the verification of the conditions (5.14)–(5.16) was discussed in Cramer and Rüschendorf (1998).

5.3 Lyapunov exponents

The classical approach to study recursion (5.2) in the affine case $K=1$ is based on properties of Lyapunov exponents. For $K=1$ the process (L^n) can be defined pointwise by

$$L^n := A_nL^{n-1}+b_n \quad \text{for all } n\ge 1, \tag{5.67}$$

where (A_n,b_n) is an i.i.d. sequence of random $n\times n$ matrices A_n and random translations b_n. Furthermore an initial r.v. L^0 is given. Iterating (5.67) L^n has the representation

$$L^n = A_n\cdot\ldots\cdot A_1L^0+\sum_{j=2}^{n}(A_n\cdot\ldots\cdot A_j)b_{j-1}+b_n. \tag{5.68}$$

The distributional asymptotics of L^n are usually analyzed introducing a *change of time* (see e.g. Verwaat (1979)): Let $Y^0 := 0$ and

$$Y^n := b_1+\sum_{j=1}^{n-1}(A_1\cdot\ldots\cdot A_j)b_{j+1}. \tag{5.69}$$

Then the distributional relation

$$L^n \stackrel{\mathcal{D}}{=} Y^n + A_1 \cdot \ldots \cdot A_n L^0 \tag{5.70}$$

holds. Assuming appropriate assumptions involving the notion of a Lyapunov exponent $A_1 \cdot \ldots \cdot A_n L^0$ becomes asymptotically small and Y^n converges a.s. to

$$Y := b_1 + \sum_{j=1}^{\infty} (A_1 \cdot \ldots \cdot A_j) b_{j+1}. \tag{5.71}$$

Following this line $L^n \to Y$ in distribution can be deduced.

The *(top) Lyapunov exponent* of a random $n \times n$ matrix A satisfying the condition $\mathbb{E} \ln^+ \|A\| < \infty$ is defined by

$$\gamma(A) := \inf_{n \in \mathbb{N}} \frac{1}{n} \mathbb{E}\,[\ln \|A_1 \cdot \ldots \cdot A_n\|], \tag{5.72}$$

where (A_i) are independent with $A_i \sim A$. The analysis of Y^n is based on the fact that

$$\gamma(A) = \lim_{n \to \infty} \frac{1}{n} \ln \|A_1 \cdot \ldots \cdot A_n\| \quad \text{a.s.} \tag{5.73}$$

which was proved in Furstenberg and Kesten (1960) and is a consequence of the subadditive ergodic theorem of Kingman (1973). If $\mathbb{E} \ln \|b\| < \infty$ and $\gamma(A) < 0$ then convergence of L^n to Y was shown in Burton and Rösler (1995).

For a scaled version of such a result consider

$$\widetilde{L}^n := V_n L^n, \quad \text{for all} \quad n \geq 1 \tag{5.74}$$

with V_n given by (5.6). Define

$$\beta := \frac{1}{2} \min_{1 \leq i \leq d} \liminf_{n \to \infty} \frac{1}{n} \ln \Big(\mathrm{Var}(L_i^n)\Big). \tag{5.75}$$

Corresponding to Theorem 5.1.3 we derive a limit law for $(\widetilde{L}^n)$ based on Lyapunov exponents in the case $K = 1$:

Theorem 5.3.1 *Let (L^n) be given by (5.67) with $\mathbb{E} \ln^+ \|b\| < \infty$, $\|L^0\| < \infty$ a.s., and $\gamma(A) < \beta$, where $\beta > 0$ is given by (5.75). Then*

$$Z := \lim_{n \to \infty} V_n \left(b_1 + \sum_{j=1}^{n-1} (A_1 \cdot \ldots \cdot A_j) b_{j+1} \right) \tag{5.76}$$

exists almost surely and

$$\widetilde{L}^n \xrightarrow{\mathcal{D}} Z \quad \text{for } n \to \infty. \tag{5.77}$$

Proof: By (5.69) and (5.70) it holds

$$\widetilde{L}^n \stackrel{\mathcal{D}}{=} V_n Y^n + V_n A_1 \cdot \ldots \cdot A_n L^0. \tag{5.78}$$

Since $\beta > 0$ and $\gamma(A) < \beta$ there exists $0 < \xi < (\beta - \gamma(A))/4$ with $\beta_- := \beta - \xi > 0$ and $\alpha_{++} := \gamma(A) + 2\xi \neq 0$. Define $\alpha_+ := \gamma(A) + \xi$. By (5.73) it exists a.s. a $n_0 = n_0(\omega) \in \mathbb{N}$ with

$$\|A_1 \cdot \ldots \cdot A_n\| \leq \exp(n\alpha_+) \quad \text{for all} \quad n \geq n_0. \tag{5.79}$$

Therefore a.s. a constant $C_1 = C_1(\omega) > 0$ exists with

$$\|A_1 \cdot \ldots \cdot A_n\| \leq C_1 \exp(n\alpha_+) \quad \text{for all} \quad n \in \mathbb{N}. \tag{5.80}$$

Analogously using (5.75) there exists a $C_2 > 0$ with

$$\|V_n\| \leq C_2 \exp(-n\beta_-) \quad \text{for all} \quad n \in \mathbb{N}. \tag{5.81}$$

Furthermore it exists a $C_3 > 0$ so that a.s.

$$\|b_n\| \leq C_3 \exp(n\xi/4) \quad \text{for all} \quad n \in \mathbb{N}. \tag{5.82}$$

This can be seen by a standard argument:

$$\begin{aligned} \sum_{n\geq 1} \mathbb{P}\Big(\|b_n\| > \exp(n\xi/4)\Big) &= \sum_{n\geq 1} \mathbb{P}\Big((4/\xi)\ln\|b_n\| > n\Big) \\ &\leq \frac{4}{\xi}\,\mathbb{E}\,\ln^+\|b\| < \infty. \end{aligned} \tag{5.83}$$

Then by the Borel-Cantelli Lemma

$$\mathbb{P}\left(\limsup_{n\to\infty}\Big\{\|b_n\| > \exp(n\xi/4)\Big\}\right) = 0. \tag{5.84}$$

This means that with probability one $\|b_n\| > \exp(n\xi/4)$ occurs only finitely many times. This implies (5.82).

Now we complete the proof showing

$$V_n A_1 \cdot \ldots \cdot A_n L^0 \longrightarrow 0 \quad \text{for} \quad n \to \infty \quad \text{a.s.}, \tag{5.85}$$

$$V_n Y^n \longrightarrow Z \quad \text{for} \quad n \to \infty \quad \text{a.s.} \tag{5.86}$$

`Ad (5.85):` Almost surely holds

$$\begin{aligned} \|V_n A_1 \cdot \ldots \cdot A_n L^0\| &\leq \|V_n\|\,\|A_1 \cdot \ldots \cdot A_n\|\,\|L^0\| \\ &\leq C_1 C_2 \exp(-n\beta_-)\exp(n\alpha_+)\|L^0\| \\ &\leq C_1 C_2 \|L^0\| \exp(-2\xi n) \to 0 \quad \text{for} \quad n \to \infty, \end{aligned} \tag{5.87}$$

by (5.80), (5.81), and $\|L^0\| < \infty$ a.s.

`Ad (5.86):` It is

$$V_nY_n = V_n \left(b_1 + \sum_{j=1}^{n-1} A_1 \cdot \ldots \cdot A_j b_{j+1} \right). \tag{5.88}$$

For the convergence of V_nY_n we show that (V_nY_n) is a Cauchy sequence a.s. We prove a.s.

$$\lim_{n_0\to\infty} \sup_{n_0\le m\le n} \|V_nY_n - V_mY_m\| = 0. \tag{5.89}$$

For $n_0 \le m < n$ it is

$$\begin{aligned}
&\|V_nY_n - V_mY_m\| \\
&= \left\| V_n \sum_{j=m}^{n-1} A_1 \cdot \ldots \cdot A_j b_{j+1} + (V_n - V_m) \sum_{j=1}^{m-1} A_1 \cdot \ldots \cdot A_j b_{j+1} \right\| \\
&\le \|V_n\| \sum_{j=m}^{n-1} \|A_1 \cdot \ldots \cdot A_j\| \, \|b_{j+1}\| \\
&\quad + \|V_n - V_m\| \sum_{j=1}^{m-1} \|A_1 \cdot \ldots \cdot A_j\| \, \|b_{j+1}\|.
\end{aligned} \tag{5.90}$$

The first summand in (5.90) denoting $C := C_1C_2C_3$ is estimated a.s. by

$$\begin{aligned}
&\|V_n\| \sum_{j=m}^{n-1} \|A_1 \cdot \ldots \cdot A_j\| \, \|b_{j+1} \\
&\le C \exp(-n\beta_-) \sum_{j=m}^{n-1} \exp(j\alpha_{++}) \\
&= C \exp(-n\beta_-) \frac{\exp(\alpha_{++})^m - \exp(\alpha_{++})^n}{1 - \exp(\alpha_{++})} \\
&= \frac{C}{1 - \exp(\alpha_{++})} \Big[\exp(m\alpha_{++} - n\beta_-) - \exp(n(\alpha_{++} - \beta_-)) \Big] \\
&\longrightarrow 0 \quad \text{for} \quad n_0 \to \infty,
\end{aligned} \tag{5.91}$$

since $\exp(m\alpha_{++} - n\beta_-) \le \exp(-n_0\xi)$ and $\exp(n(\alpha_{++} - \beta_-)) \le \exp(-n_0\xi)$. For the second summand of (5.90) observe

$$\|V_n - V_m\| \le C_2 \exp(-m\beta_-), \tag{5.92}$$

since $m < n$ and $\beta_- > 0$. This implies a.s.

$$
\begin{aligned}
\|V_n - V_m\| \sum_{j=1}^{m-1} \|A_1 \cdot \ldots \cdot A_j\| \, \|b_{j+1}\| & \\
\leq \quad & C \exp(-m\beta_-) \sum_{j=1}^{m-1} \exp(j\alpha_{++}) \\
= \quad & C \exp(-m\beta_-) \frac{1 - \exp(\alpha_{++})^m}{1 - \exp(\alpha_{++})} \\
= \quad & \frac{C}{1 - \exp(\alpha_{++})} \Big[\exp(-m\beta_-) - \exp(m(\alpha_{++} - \beta_-)) \Big] \\
\longrightarrow \quad & 0 \quad \text{for} \quad n_0 \to \infty,
\end{aligned} \tag{5.93}
$$

since $\exp(-m\beta_-) \leq \exp(-n_0\xi)$ and $\exp(n(\alpha_{++} - \beta_-)) \leq \exp(-n_0\xi)$. ■

Even in the unscaled case with $K = 1$ there is not much known about the connection of the conditions arising in the L_2-case formulated in expectations of norms of certain matrices as in (5.29) with conditions on Lyapunov exponents. For a discussion of this problem and a related conjecture see Burton and Rösler (1995).

Bibliography

Adler, R. L. and L. Flatto (1977). Uniform distribution of Kakutani's interval splitting procedure. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **38**, 253–259.

Aldous, D. (1996). Probability distributions on cladograms. In D. Aldous and R. Permantle (Eds.), *Random Discrete Structures*, Volume 76 of *IMA Volumes Math. Appl.*, pp. 1–18. Springer.

Alsmeyer, G. (1991). Complete answer to an interval splitting problem. *Statistics & Probability Letters* **12**, 285–287.

Anderson, D. H. and R. Brown (1992). Combinatorial aspects of C.A.R. Hoare's FIND algorithm. *Australasian Journal of Combinatorics* **5**, 109–119.

Arora, S. and W. Dent (1969). Randomized binary search technique. *Communications of the ACM* **12**, 77–80.

Assche, W. van (1986). Products of 2×2 stochastic matrices with random entries. *Journal of Applied Probability* **23**, 1019–1024.

Assche, W. van (1987). A random variable uniformly distributed between two independent random variables. *Sankhyã: The Indian Journal of Statistics* **49,** Series A, 207–211.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM* **18**, 509–517.

Bougerol, P. and N. Picard (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability* **20**, 1714–1730.

Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* **18**, 211–220.

Bruhn, V. (1996). *Eine Methode zur asymptotischen Behandlung einer Klasse von Rekursionen mit einer Anwendung in der stochastischen Analyse des Quicksort-Algorithmus*. Ph. D. thesis, Christian-Albrechts-Universität zu Kiel.

Bruss, T., R. Jammalamadaka, and X. Zhou (1990). On an interval splitting problem. *Statistics & Probability Letters* **10**, 321–324.

Burton, R. and U. Rösler (1995). An L_2 convergence theorem for random affine mappings. *Journal of Applied Probability* **32**, 183–192.

Chamayou, J.-F. and G. Letac (1991). Explicit stationary distributions for compositions of random functions and products of random matrices. *Journal of Theoretical Probability* **4**, 3–36.

Chamayou, J.-F. and G. Letac (1994). A transient random walk on stochastic matrices with Dirichlet distribution. *The Annals of Probability* **22**, 424–430.

Chen, R., R. Goodman, and A. Zame (1984). Limiting distributions of two random sequences. *Journal of Multivariate Analysis* **14**, 221–230.

Chen, R., E. Lin, and A. Zame (1981). Another acr sine law. *Sankhyã: The Indian Journal of Statistics* **43,** Series A, 371–373.

Cramer, M. (1995). *Stochastische Analyse rekursiver Algorithmen mit idealen Metriken*. Ph. D. thesis, Albert-Ludwigs-Universität Freiburg.

Cramer, M. and L. Rüschendorf (1996). Convergence of a branching type recursion. *Annales de l'Institut Henri Poincaré* **32**, 725–741.

Cramer, M. and L. Rüschendorf (1998). Convergence of two-dimensional branching recursions. Preprint.

Cunto, W., G. Lau, and P. Flajolet (1989). Analysis of kdt–trees: kd–trees improved by local reorganisations. In F. Dehne, J.-R. Sack, and N. Santoro (Eds.), *Algorithms and Data Structures*, Volume 382 of *LNCS*, pp. 24–38.

Devroye, L. (1986). A note on the expected height of binary search trees. *Journal of the ACM* **33**, 489–498.

Devroye, L. (1998). Universal limit laws for the depths in random trees. *Journal on Computing* **28**, 409–432.

Devroye, L., J. Jabbour, and C. Zamora-Cura (1999). Squarish k-d trees. Preprint.

Devroye, L. and L. Laforest (1990). An analysis of random d–dimensional quad trees. *SIAM Journal on Computing* **19**, 821–832.

Devroye, L., G. Letac, and V. Seshadri (1986). The limit behavior of an interval splitting scheme. *Statistics & Probability Letters* **4**, 183–186.

Dobrow, R. P. and J. A. Fill (1999). Total path length for random recursive trees. To appear in *Combinatorics, Probability and Computing.*

Duch, A., V. Estivill-Castro, and C. Martínez (1998). Randomized k-dimensional binary search trees. Preprint.

Durrett, R. and T. M. Liggett (1983). Fixed points of the smoothing transformation. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**, 275–301.

Finkel, R. and J. Bentley (1974). Quad trees, a data structure for retrieval on composite keys. *Acta informatica* **4**, 1–9.

Flajolet, P., G. Gonnet, C. Puech, and J. Robson (1993). Analytic variations on quadtrees. *Algorithmica* **10**, 473–500.

Flajolet, P., G. Labelle, L. Laforest, and B. Salvy (1995). Hypergeometrics and the cost structure of quadtrees. *Random Structures and Algorithms* **7**, 117–144.

Flajolet, P. and C. Puech (1986). Partial match retrieval of multidimensional data. *Journal of the ACM* **33**, 371–407.

Furstenberg, H. and H. Kesten (1960). Products of random matrices. *Annals of Mathematical Statistics* **31**, 457–469.

Grübel, R. (1998). Hoare's selection algorithm: A Markov chain approach. *Journal of Applied Probability* **35**, 36–45.

Grübel, R. (1999). On the median-of-k version of Hoare's selection algorithm. *Theoretical Informatics and Applications* **33**, 177–192.

Grübel, R. and U. Rösler (1996). Asymptotic distribution theory for Hoare's selection algorithm. *Advances in Applied Probability* **28**, 252–269.

Guivarc'h, Y. (1990). Sur une extension de la notion de loi semi-stable. *Annales de l'Institut Henri Poincaré* **26**, 261–286.

Hoare, C. A. R. (1961). Algorithm 64, Quicksort; Algorithm 65, Find. *Communications of the ACM* **4**, 321–322.

Hoare, C. A. R. (1962). Quicksort. *The Computer Journal* **5**, 10–15.

Holley, R. and T. M. Liggett (1981). Generalized potlatch and smoothing processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **55**, 165–195.

Jacquet, P. and W. Szpankowski (1995). Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science* **144**, 161–197.

Johnson, N. L. and S. Kotz (1990). Use of moments in deriving distributions and some characterizations. *Mathematical Scientist* **15**, 42–52.

Johnson, N. L. and S. Kotz (1995). Use of moments in studies of limit distributions arising from iterated random subdivisions of an interval. *Statistics & Probability Letters* **24**, 111–119.

Kahane, J.-P. and J. Peyrière (1976). Sur certaines martingales de Benoit Mandelbrot. *Advances in Mathematics* **22**, 131–145.

Kakutani, S. (1975). A problem of equidistribution on the unit interval $[0, 1]$. In *Proceedings of the Oberwolfach Conference on Measure Theory*, Volume 541 of *LNM*, pp. 369–375.

Kennedy, D. P. (1988). A note on stochastic search methods for global optimization. *Advances in Applied Probability* **20**, 476–478.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* **131**, 207–248.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *Annals of Probability* **1**, 883–909.

Kirschenhofer, P., C. Martínez, and H. Prodinger (1997). Analysis of Hoare's Find algorithm with median-of-three partition. *Random Structures and Algorithms* **10**, 143–156.

Kirschenhofer, P. and H. Prodinger (1998). Comparisons in Hoare's FIND algorithm. *Combinatorics, Probability and Computing* **7**, 111–120.

Knessl, C. and W. Szpankowski (1999). Quicksort algorithm again revisited. *Discrete Mathematics and Theoretical Computer Science* **3**, 43–64.

Knuth, D. E. (1972). Mathematical analysis of algorithms. In *Information processing 71*, Volume 10, pp. 19–27. North Holland Publishing Company 1972. Proceedings of IFIP Congress Ljubljana, 1971.

Knuth, D. E. (1997a). *The Art of Computer Programming: Fundamental Algorithms* (third ed.), Volume 1. Addison Wesley.

Knuth, D. E. (1997b). *The Art of Computer Programming: Seminumerical Algorithms* (third ed.), Volume 2. Addison Wesley.

Knuth, D. E. (1998). *The Art of Computer Programming: Sorting and Searching* (second ed.), Volume 3. Addison Wesley.

Lent, J. and H. Mahmoud (1996). Average-case analysis of multiple quickselect: An algorithm for finding order statistics. *Statistics & Probability Letters* **28**, 299–310.

Letac, G. and M. Scarsini (1998). Random nested tetrahedra. *Advances in Applied Probability* **30**, 619–627.

Mahmoud, H. (1986). On the average internal path length of m-ary search trees. *Acta Informatica* **23**, 111–117.

Mahmoud, H. (1992). *Evolution of Random Search Trees*. John Wiley.

Mahmoud, H., R. Modarres, and R. Smythe (1995). Analysis of quickselect: An algorithm for order statistics. *RAIRO, Theoretical Informatics and Applications* **28**, 299–310.

Mahmoud, H. and R. Smythe (1998). Probabilistic analysis of MULTIPLE QUICKSELECT. *Algorithmica* **22**, 569–584.

Mandelbrot, B. (1974). Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire. *C.R. Acad. Sc. Paris* **278**, Série A, 289–292.

Martínez, C., A. Panholzer, and H. Prodinger (1998). Partial match queries in relaxed multidimensional search trees. Preprint.

Martínez, C. and S. Roura (1998). Optimal sampling strategies in Quicksort and Quickselect. Preprint.

McDiarmid, C. J. H. and R. B. Hayward (1996). Large deviations for quicksort. *Journal of Algorithms* **21**, 476–507.

Mountford, S. T. and S. C. Port (1993). Random splittings of an interval. *Journal of Applied Probability* **30**, 131–152.

Neininger, R. and L. Rüschendorf (1999). On the internal path length of d-dimensional quad trees. *Random Structures and Algorithms* **15**, 25–41.

Panholzer, A. (1997). *Untersuchungen zur durchschnittlichen Gestalt gewisser Baumfamilien. Mit besonderer Berücksichtigung von Anwendungen in der Informatik*. Ph. D. thesis, Technischen Universität Wien.

Panholzer, A. and H. Prodinger (1998). A generating functions approach for the analysis of grand averages for Multiple QUICKSELECT. *Random Structures and Algorithms* **13**, 189–209.

Paulsen, V. (1997). The moments of Find. *Journal of Applied Probability* **34**, 1079–1082.

Prodinger, H. (1995). Multiple quickselect — Hoare's FIND algorithm for several elements. *Information Processing Letters* **56**, 123–129.

Pyke, R. (1980). The asymptotic behavior of spacings under Kakutani's model for interval subdivision. *The Annals of Probability* **8**, 157–163.

Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models.* John Wiley.

Rachev, S. T. and L. Rüschendorf (1995). Probability metrics and recursive algorithms. *Advances in Applied Probability* **27**, 770–799.

Régnier, M. (1989). A limiting distribution for quicksort. *RAIRO, Theoretical Informatics and Applications* **23**, 335–343.

Rosenblatt, M. (1964). Equicontinuous Markov operators. *Theory of Probability and its Applications* **9**, 180–197.

Rösler, U. (1991). A limit theorem for "quicksort". *RAIRO, Theoretical Informatics and Applications* **25**, 85–100.

Rösler, U. (1992). A fixed point theorem for distributions. *Stochastic Processes and Applications* **42**, 195–214.

Rösler, U. (1998). A fixed point equation for distributions. Preprint.

Rösler, U. (1999). On the analysis of stochastic divide and conquer algorithms. To appear in *Algorithmica*.

Rösler, U. and L. Rüschendorf (1999). The contraction method for recursive algorithms. To appear in *Algorithmica*.

Samet, H. (1990). *The Design and Analysis of Spatial Data Structures*. Addison-Wesley.

Schachinger, W. (1999). Limiting distributions for the costs of partial match retrievals in multidimensional tries. Preprint.

Slud, E. (1978). Entropy and maximal spacings for random partitions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **41**, 341–352.

Sun, T.-C. (1975). Limits of convolutions of probability measures on the set of 2×2 stochastic matrices. *Bulletin of the Institute of Mathematics Academia Sinica* **3**, 235–248.

Tan, K. H. and P. Hadjicostas (1995). Some properties of a limiting distribution in quicksort. *Statistics & Probability Letters* **25**, 87–94.

Verwaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitly divisible random variables. *Advances in Applied Probability* **11**, 750–783.

Volodin, N. A., S. Kotz, and N. L. Johnson (1993). Use of moments in distribution theory: A mutivariate case. *Journal of Multivariate Analysis* **46**, 112–119.

Zwet, W. R. van (1978). A proof of Kakutani's conjecture on random subdivision of longest intervals. *The Annals of Probability* **6**, 133–137.