Limit Laws for Partial Match Queries in Quadtrees

RALPH NEININGER AND LUDGER RÜSCHENDORF
Universität Freiburg

It is proved that in the uniform probabilistic model the cost of a partial match query in a multidimensional quadtree after normalization converges in distribution. The limiting distribution is given as a fixed point of a random affine operator. Also a first order asymptotic for the variance of the cost is derived and results on exponential moments are given. The analysis is based on the contraction method.

Key words: quadtrees, partial match query, contraction method, multidimensional data structure, analysis of algorithms.

AMS subject classification: Primary 68Q25, 60F05; secondary 68P10.

1 Introduction

A partial match query is one of several types of queries in a file which maintains multidimensional data. Databases for multidimensional data are of special interest for applications in geographical information systems, computer graphics and computational geometry. 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. Data structures which support associative queries 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* and *digital search trees*. 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 an analysis of the performance of basic parameters for these structures see Mahmoud (1992).

In this paper we give an asymptotic probabilistic analysis of the cost of partial match queries in quadtrees. 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 $q_i \neq *$. 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.

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. For a partial match query in a quadtree 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 in a quadtree is measured by the number of nodes traversed during this search. We denote this cost in a quadtree containing n nodes by C_n.

The cost has already been studied in the uniform probabilistic model. In dimension $d = 2$ and with $s = 1$ component specified the first order asymptotics for the mean and variance are known. Flajolet, Gonnet, Puech and Robson (1991, 1993) derive

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

with

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

Martínez, Panholzer and Prodinger (1998) recently found

$$\mathrm{Var}(C_n) \sim \beta n^{2\alpha-2} \tag{3}$$

with

$$\beta = \frac{(2\alpha-1)\Gamma(2\alpha)}{3\alpha(\alpha-1)\Gamma^4(\alpha)} - \frac{\Gamma^2(2\alpha)}{4\Gamma^6(\alpha)}. \tag{4}$$

In arbitrary dimension d with $1 \leq s \leq d-1$ components specified Flajolet et al. (1993) derived

$$\mathbb{E}C_n \sim \gamma_{s,d} n^{\alpha-1} \tag{5}$$

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

$$\alpha^{d-s}(\alpha+1)^s = 2^d. \tag{6}$$

An asymptotic for the variance of C_n has not been known so far.

In this paper we give limit laws for C_n in any dimension and derive the first order asymptotic of the variance of C_n and results on exponential moments. The normalized cost

$$X_n = \frac{C_n - \mathbb{E}C_n}{n^{\alpha-1}} \tag{7}$$

converges weakly to a random variable which is characterized as the fixed point of a random affine operator. For the proof we use the contraction method. This method has been introduced by Rösler (1991) for the analysis of Quicksort. The contraction method has been further developed independently in Rösler (1992) and Rachev and Rüschendorf (1995). For a recent survey of this method see also Rösler and Rüschendorf (1999).

Limit laws for the cost of a partial match queries in the uniform probabilistic modell for the K-d tries and some sorts of K-d trees have been derived recently in Schachinger (1999) respectively Neininger (1999). In all these data structures the mean and standard deviation for the cost of a partial match query are known to be of the same order of magnitude. Therefore we do not need the second order term in the expansion of the mean of C_n in order to derive a limiting operator. The second order term for the mean has turned out to be crucial for the problem of the internal path length of related random trees (see Dobrow and Fill (1999), Rösler (1999) and Neininger and Rüschendorf (1999)). From this point of view the problem of partial match query contains some similarity with the running time of the FIND-algorithm in the model of Mahmoud, Modarres and Smythe (1995). Nevertheless for the FIND problem it is easier to derive information on the limit distribution from the fixed point equation. This is caused by the purely one-sided character of the FIND-algorithm.

The authors thank P. Flajolet for his exposure of the problem of a partial match query and for pointing out its connection to the contraction method during a visit in Freiburg.

2 Standard Quadtrees in Dimension d=2

Denote by $W = (U, V)$ the first key to be inserted, which is stored in the root of the random 2-dimensional Quadtree. U, V are independent and uniformly distributed on $[0, 1]$. The root $W = w = (u, v)$ partitions the unit square into 4 quadrants with volumes given by $\langle w \rangle := (uv, u(1-v), (1-u)v, (1-u)(1-v))$. Denote by $I^{(n)}$ the vector of the cardinalities of the subtrees of the root of a random quadtree with n nodes. Then conditionally given $W = w$ the vector $I^{(n)}$ is multinomial $M(n-1, \langle w \rangle)$ distributed:

$$\mathbb{P}^{I^{(n)}|W=w} = M(n-1, \langle w \rangle). \tag{8}$$

The weak law of large numbers for $I^{(n)}$ then implies:

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle W \rangle = (UV, U(1-V), (1-U)V, (1-U)(1-V)) \tag{9}$$

with (U, V) uniformly distributed on $[0, 1]^2$.
This implies L^1-convergence of bounded continuous functionals of $I^{(n)}/n$, in

particular

$$\mathbb{E}\left(\frac{I_k^{(n)}}{n}\right)^{2\alpha-2} \longrightarrow \mathbb{E}(UV)^{2\alpha-2} = \frac{1}{(2\alpha-1)^2}, \tag{10}$$

$$\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\left(\frac{I_1^{(n)}}{n}\right)^{2\alpha-2}\right] \longrightarrow \mathbb{E}[\mathbf{1}_{\{Y<U\}}(UV)^{2\alpha-2}] = \frac{1}{2\alpha(2\alpha-1)} \tag{11}$$

if Y is uniformly distributed on $[0,1]$ and independent of $I^{(n)}, U$.

For a partial match query in dimension 2 one component of the search pattern is specified. W.l.g. we can assume the first component is specified, so the pattern is of the form $(S,*)$. For the distributional analysis of partial match query let C_n denote the number of nodes traversed in the quadtree during a partial match retrieval. We assume that the first component of the search pattern Y is uniformly distributed on $[0,1]$ and independent of the random quadtree according to the uniform probabilistic model. Then $C_1 = 1$; we define $C_0 := 0$. Conditionally given $I^{(n)}$ the subtrees are mutually independent and distributed as quadtrees. For this reason the number of traversed nodes C_n satisfies the distributional recursive equation

$$C_n \stackrel{\mathcal{D}}{=} \mathbf{1}_{\{Y<U\}}\left(C^{(1)}_{I_1^{(n)}} + C^{(2)}_{I_2^{(n)}}\right) + \mathbf{1}_{\{Y\geq U\}}\left(C^{(3)}_{I_3^{(n)}} + C^{(4)}_{I_4^{(n)}}\right) + 1. \tag{12}$$

Here $Y, U, V, (C_i^{(1)}), \dots, (C_i^{(4)})$ are independent, Y, U, V are uniformly distributed on $[0,1]$, $C_i^{(k)} \stackrel{\mathcal{D}}{=} C_i$ for $k = 1, \dots, 4$ and $I^{(n)}$ is given $(U,V) = w$ multinomial $M(n-1, \langle w \rangle)$ distributed. Related independence properties are used throughout the paper in recursive equations of a similar form without stating them explicitly in each case.

Recall the first order asymptotics for the mean and variance of C_n. Flajolet, Gonnet, Puech and Robson (1993) derive

$$\mathbb{E}C_n \sim \gamma n^{\alpha-1} \tag{13}$$

with α and γ given in (2).

In Martínez, Panholzer and Prodinger (1998) it is shown that

$$\mathrm{Var}(C_n) \sim \beta n^{2\alpha-2} \tag{14}$$

with β as in (4).

Therefore a normalized versionX_n of C_n is given by

$$X_n := \frac{C_n - \mathbb{E}C_n}{n^{\alpha-1}}. \tag{15}$$

X_n satisfies the modified recursion

$$\begin{aligned} X_n \stackrel{\mathcal{D}}{=} & \quad \mathbf{1}_{\{Y<U\}} \left[\sum_{k=1}^{2} \left(\frac{I_k^{(n)}}{n} \right)^{\alpha-1} \left(X_{I_k^{(n)}}^{(k)} + \gamma + o(1) \right) \right] \\ & + \quad \mathbf{1}_{\{Y\geq U\}} \left[\sum_{k=3}^{4} \left(\frac{I_k^{(n)}}{n} \right)^{\alpha-1} \left(X_{I_k^{(n)}}^{(k)} + \gamma + o(1) \right) \right] - \gamma + o(1). \end{aligned} \tag{16}$$

This modified recursion and (9) suggest that a limit X of X_n is a solution of the limiting equation

$$\begin{aligned} X \stackrel{\mathcal{D}}{=} & \quad \mathbf{1}_{\{Y<U\}} \left[(UV)^{\alpha-1}(X^{(1)} + \gamma) + (U(1-V))^{\alpha-1}(X^{(2)} + \gamma) \right] \\ & + \quad \mathbf{1}_{\{Y\geq U\}} \left[((1-U)V)^{\alpha-1}(X^{(3)} + \gamma) \right. \\ & \qquad \left. + ((1-U)(1-V))^{\alpha-1}(X^{(4)} + \gamma) \right] - \gamma. \end{aligned} \tag{17}$$

Here $Y, U, V, X^{(1)}, \ldots, X^{(4)}$ are independent, Y, U, V are uniformly distributed on $[0,1]$ and $X^{(k)} \stackrel{\mathcal{D}}{=} X$ for $k = 1, \ldots, 4$.

Define

$$M_{0,2} := \{ \mu \in M^1(\mathbb{R}^1, \mathcal{B}^1) : \mathbb{E}\mu = 0, \ \mathrm{Var}\,\mu < \infty \} \tag{18}$$

where $\mathbb{E}\mu, \mathrm{Var}\,\mu$ are defined as expectation respectively variance of a corresponding random variable and $M^1(\mathbb{R}^1, \mathcal{B}^1)$ denotes the space of probability measures on the real line. Define the random affine operator corresponding to (17)

$$\begin{aligned} & T : M^1(\mathbb{R}^1, \mathcal{B}^1) \to M^1(\mathbb{R}^1, \mathcal{B}^1) \\ & T(\mu) \stackrel{\mathcal{D}}{=} \mathbf{1}_{\{Y<U\}} \left[(UV)^{\alpha-1}(Z^{(1)} + \gamma) + (U(1-V))^{\alpha-1}(Z^{(2)} + \gamma) \right] \\ & \qquad + \mathbf{1}_{\{Y\geq U\}} \left[((1-U)V)^{\alpha-1}(Z^{(3)} + \gamma) \right. \\ & \qquad\qquad \left. + ((1-U)(1-V))^{\alpha-1}(Z^{(4)} + \gamma) \right] - \gamma, \end{aligned} \tag{19}$$

where $Y, U, V, Z^{(1)}, \dots, Z^{(4)}$ are independent, Y, U, V are uniformly distributed on $[0,1]$, $Z^{(k)} \stackrel{\mathcal{D}}{=} \mu$ for $k = 1, \dots, 4$.

Our aim is to show that T is the *limiting operator* of the recursive sequence (X_n) in (16). Supply $M_{0,2} \subset M^1(\mathbb{R}^1, \mathcal{B}^1)$ with the minimal l_2–metric

$$l_2(\mu, \nu) = \inf\{(\mathbb{E}|X - Y|^2)^{1/2} : X \stackrel{\mathcal{D}}{=} \mu, Y \stackrel{\mathcal{D}}{=} \nu\}. \tag{20}$$

For random variables X, Y we use synonymously $l_2(X, Y) = l_2(\mathbb{P}^X, \mathbb{P}^Y)$. Then $(M_{0,2}, l_2)$ is a complete metric space and $l_2(\mu_n, \mu) \to 0$ is equivalent to

$$\mu_n \stackrel{\mathcal{D}}{\to} \mu \quad \text{and} \quad \int x^2 \, \mathrm{d}\mu_n(x) \to \int x^2 \, \mathrm{d}\mu(x). \tag{21}$$

(see Rachev (1991))

Lemma 2.1 *$T : M_{0,2} \to M_{0,2}$, with T given in (19) is a contraction w.r.t. l_2:*

$$l_2(T(\mu), T(\nu)) \le \xi \, l_2(\mu, \nu) \quad \text{for all} \quad \mu, \nu \in M_{0,2}, \tag{22}$$

$$\xi = \frac{2}{\sqrt{19 - 3\sqrt{17}}} = 0.776\dots \tag{23}$$

Proof: Obviously $\mathrm{Var}\ T(\mu) < \infty$. $\mathbb{E}T(\mu) = 0$ follows from $\mathbb{E}[\mathbf{1}_{\{Y<U\}} U^{\alpha-1} V^{\alpha-1}] = 1/4$. 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 $(W^{(k)}, Z^{(k)}), Y, U, V$ be independent, Y, U, V uniformly distributed on $[0,1]$. Let $(W^{(k)}, Z^{(k)})$ be optimal l_2–couplings of (μ, ν), i.e. $W^{(k)} \stackrel{\mathcal{D}}{=} \mu$, $Z^{(k)} \stackrel{\mathcal{D}}{=} \nu$ and $l_2^2(\mu, \nu) = \mathbb{E}(W^{(k)} - Z^{(k)})^2$ for $k = 1, \dots, 4$. Then using the independence properties and $\mathbb{E}W^{(k)} = \mathbb{E}Z^{(k)} = 0$

$$\begin{aligned}
&l_2^2(T(\mu), T(\nu)) \\
&\le \mathbb{E}\Big|\mathbf{1}_{\{Y<U\}}\Big\{(UV)^{\alpha-1}(W^{(1)} - Z^{(1)}) + (U(1-V))^{\alpha-1}(W^{(2)} - Z^{(2)})\Big\} \\
&\qquad + \mathbf{1}_{\{Y \ge U\}}\Big\{((1-U)V)^{\alpha-1}(W^{(3)} - Z^{(3)}) \\
&\qquad\qquad + ((1-U)(1-V))^{\alpha-1}(W^{(4)} - Z^{(4)})\Big\}\Big|^2 \\
&= \mathbb{E}\Big[\mathbf{1}_{\{Y<U\}}\Big\{(UV)^{2\alpha-2}(W^{(1)} - Z^{(1)})^2 + (U(1-V))^{2\alpha-2}(W^{(2)} - Z^{(2)})^2\Big\}
\end{aligned}$$

$$
\begin{aligned}
&+\mathbf{1}_{\{Y \geq U\}} \Big\{((1-U)V)^{2\alpha-2}(W^{(3)}-Z^{(3)})^2 \\
&\qquad +((1-U)(1-V))^{2\alpha-2}(W^{(4)}-Z^{(4)})^2\Big\}\Big] \\
&= 4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}(UV)^{2\alpha-2}(W^{(1)}-Z^{(1)})^2\right] \\
&= 4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}(UV)^{2\alpha-2}\right] l_2^2(\mu,\nu). \qquad (24)
\end{aligned}
$$

From

$$
\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}(UV)^{2\alpha-2}\right] = \frac{1}{2\alpha(2\alpha-1)} = \frac{1}{19-3\sqrt{17}}. \qquad (25)
$$

the assertion follows. ■

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

$$
l_2(T^n(\mu),\rho) \to 0 \qquad (26)
$$

exponentially fast for any $\mu \in M_{0,2}$.

We call a random variable X with distribution ρ also a fixed point of T. (compare equation (17))

The representation of the limiting operator T can be simplified. We have

$$
T(\mu) \stackrel{\mathcal{D}}{=} U^{\frac{\alpha-1}{2}}\{V^{\alpha-1}(Z^{(1)}+\gamma)+(1-V)^{\alpha-1}(Z^{(2)}+\gamma\})-\gamma \qquad (27)
$$

with $U, V, Z^{(1)}, Z^{(2)}$ being independent, U, V uniformly distributed on $[0,1]$ and $Z^{(1)}, Z^{(2)} \stackrel{\mathcal{D}}{=} \mu$. The proof follows from an elementary calculation observing that the sets of the indicator function in (19) 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 $\tilde{X} := X + \gamma$ is a fixed point of

$$
\tilde{T}(\mu) :\stackrel{\mathcal{D}}{=} U^{\frac{\alpha-1}{2}}(V^{\alpha-1}\tilde{Z}^{(1)}+(1-V)^{\alpha-1}\tilde{Z}^{(2)}) \qquad (28)
$$

in $M_{\gamma,2} := \{\mu \in M^1(\mathbb{R}^1, \mathcal{B}^1) \mid \mathbb{E}\mu = \gamma,\ \mathrm{Var}\,\mu < \infty\}$.

Theorem 2.2 (Limit Theorem for Partial Match Query in 2-dimensional Quadtrees) *The normalized number of nodes traversed during a partial match* *query in a random 2-dimensional quadtree X_n converges w.r.t. l_2 to the unique fixed point X in $M_{0,2}$ of the limiting operator T, i.e.*

$$l_2(X_n, X) \to 0. \tag{29}$$

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

$$Z \stackrel{\mathcal{D}}{=} U^{\frac{\alpha-1}{2}}(V^{\alpha-1}Z^{(1)} + (1-V)^{\alpha-1}Z^{(2)}) \tag{30}$$

with $U, V, Z^{(1)}, Z^{(2)}$ independent, U, V uniformly distributed on $[0,1]$ and $Z^{(1)}, Z^{(2)} \stackrel{\mathcal{D}}{=} Z$.

Proof: Let $X_n^{(k)} \stackrel{\mathcal{D}}{=} X_n, X^{(k)} \stackrel{\mathcal{D}}{=} X$ for $k = 1, \dots, 4$ such that $(X_n^{(k)}, X^{(k)})$ are optimal couplings of X_n, X, i.e. $l_2^2(X_n, X) = \mathbb{E}(X_n^{(k)} - X^{(k)})^2$. Furthermore let $I^{(n)}$ be conditionally given $(U,V) = \langle w \rangle$ multinomial $M(n-1, \langle w \rangle)$ distributed. By (9) then $I^{(n)}/n \to \langle (U,V) \rangle$ in probability. Let U, V and Y being independent and uniformly distributed on $[0,1]$. Finally assume that $((X_n^{(1)})_{n \in \mathbb{N}}, X^{(1)}), \dots, ((X_n^{(4)})_{n \in \mathbb{N}}, X^{(4)}), (I^{(n)}, U, V), Y$ are independent.
For the estimate of $l_2(X_n, X)$ we use the L_2-distance of the special representation of X_n and X given by (16) respectively (19). Then using the independence properties and that $\mathbb{E}X^{(k)} = \mathbb{E}X_n^{(k)} = 0$ we obtain

$$\begin{aligned}
&l_2^2(X_n, X) \\
\leq\ & \mathbb{E}\Bigg[\Bigg(\mathbf{1}_{\{Y<U\}}\Bigg(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}} + \gamma\right) - (UV)^{\alpha-1}(X^{(1)} + \gamma) \\
&\quad + \left(\frac{I_2^{(n)}}{n}\right)^{\alpha-1}\left(X^{(2)}_{I_2^{(n)}} + \gamma\right) - (U(1-V))^{\alpha-1}(X^{(2)} + \gamma)\Bigg) \\
&\quad + \mathbf{1}_{\{Y \geq U\}}\Bigg(\left(\frac{I_3^{(n)}}{n}\right)^{\alpha-1}\left(X^{(3)}_{I_3^{(n)}} + \gamma\right) - ((1-U)V)^{\alpha-1}(X^{(3)} + \gamma) \\
&\quad + \left(\frac{I_4^{(n)}}{n}\right)^{\alpha-1}\left(X^{(4)}_{I_4^{(n)}} + \gamma\right) - ((1-U)(1-V))^{\alpha-1}(X^{(4)} + \gamma)\Bigg) \\
&+ o(1)\Bigg)^2\Bigg]
\end{aligned}$$

$$
\begin{aligned}
= \ & \mathbb{E}\Bigg[\mathbf{1}_{\{Y<U\}}\Bigg\{\Bigg(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}}+\gamma\right)-(UV)^{\alpha-1}(X^{(1)}+\gamma)\Bigg)^2 \\
& \quad +\Bigg(\left(\frac{I_2^{(n)}}{n}\right)^{\alpha-1}\left(X^{(2)}_{I_2^{(n)}}+\gamma\right)-(U(1-V))^{\alpha-1}(X^{(2)}+\gamma)\Bigg)^2\Bigg\} \\
& \quad +\mathbf{1}_{\{Y\geq U\}}\Bigg\{\Bigg(\left(\frac{I_3^{(n)}}{n}\right)^{\alpha-1}\left(X^{(3)}_{I_3^{(n)}}+\gamma\right)-((1-U)V)^{\alpha-1}(X^{(3)}+\gamma)\Bigg)^2 \\
& \quad +\Bigg(\left(\frac{I_4^{(n)}}{n}\right)^{\alpha-1}\left(X^{(4)}_{I_4^{(n)}}+\gamma\right)-((1-U)(1-V))^{\alpha-1}(X^{(4)}+\gamma)\Bigg)^2\Bigg\}\Bigg] \\
& +o(1), \qquad (31)
\end{aligned}
$$

where the mixed terms are $o(1)$ using independence and $\mathbb{E}[(I_1^{(n)}/n)^{\alpha-1} - (UV)^{\alpha-1}] = o(1)$ (analogously for $I_2^{(n)}, I_3^{(n)}, I_4^{(n)}$). The four occurring summands in (31) are identically distributed. This implies

$$
\begin{aligned}
& l_2^2(X_n, X) \\
\leq \ & 4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\Bigg(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}}+\gamma\right)-(UV)^{\alpha-1}(X^{(1)}+\gamma)\Bigg)^2\right] \\
& +o(1) \\
= \ & 4\mathbb{E}\Bigg[\mathbf{1}_{\{Y<U\}}\Bigg(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right) \\
& \qquad +\left(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}-(UV)^{\alpha-1}\right)(X^{(1)}+\gamma)\Bigg)^2\Bigg] + o(1) \\
= \ & 4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\left(\frac{I_1^{(n)}}{n}\right)^{2\alpha-2}\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)^2\right]
\end{aligned}
$$

$$+4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\left(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}-(UV)^{\alpha-1}\right)^2 (X^{(1)}+\gamma)^2\right]$$

$$+8\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)\right.$$

$$\left.\times\left(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}-(UV)^{\alpha-1}\right)(X^{(1)}+\gamma)\right]+o(1). \quad (32)$$

In consequence of (9) we obtain

$$\mathbb{E}\left(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}-(UV)^{\alpha-1}\right)^2 \to 0 \text{ as } n\to\infty. \quad (33)$$

Therefore the second summand of (32) converges to 0. With the Cauchy-Schwarz inequality and (33) the third term in its absolute value is estimated from above by

$$2\mathbb{E}\left[\left(\left(\frac{I_1^{(n)}}{n}\right)^{\alpha-1}-(UV)^{\alpha-1}\right)^2 (X^{(1)}+\gamma)^2\right]^{1/2} \mathbb{E}\left[\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)^2\right]^{1/2}$$

$$= \; o(1)\mathbb{E}\left[\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)^2\right]^{1/2} \leq o(1)\mathbb{E}\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)^2. \quad (34)$$

For the last inequality observe that if the expectation is less than 1 then both sides are $o(1)$. Therefore we derive from (32) with $a_n := l_2^2(X_n, X)$

$$\begin{aligned}
a_n &\leq 4\mathbb{E}\left[\left(\mathbf{1}_{\{Y<U\}}\left(\frac{I_1^{(n)}}{n}\right)^{2\alpha-2}+o(1)\right)\left(X^{(1)}_{I_1^{(n)}}-X^{(1)}\right)^2\right]+o(1)\\
&= 4\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\mathbf{1}_{\{I_1^{(n)}=i\}}\mathbf{1}_{\{Y<U\}}\left(\frac{i}{n}\right)^{2\alpha-2}+o(1)\right)\left(X_i^{(1)}-X^{(1)}\right)^2\right]+o(1)\\
&= 4\sum_{i=0}^{n-1}\mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}=i\}}\mathbf{1}_{\{Y<U\}}\left(\frac{i}{n}\right)^{2\alpha-2}+o(1)\right] l_2^2(X_i,X)+o(1). \quad (35)
\end{aligned}$$

Thus from (11)

$$\begin{aligned}
a_n &\leq 4\sum_{i=0}^{n-1} \mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}=i\}}\mathbf{1}_{\{Y<U\}}\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] \sup_{1\leq i\leq n-1} a_i + o(1) \\
&= \left(4\mathbb{E}\left[\mathbf{1}_{\{Y<U\}}(UV)^{2\alpha-2}\right] + o(1)\right) \sup_{1\leq i\leq n-1} a_i + o(1) \\
&= (\xi^2 + o(1)) \sup_{1\leq i\leq n-1} a_i + o(1) \qquad (36)
\end{aligned}$$

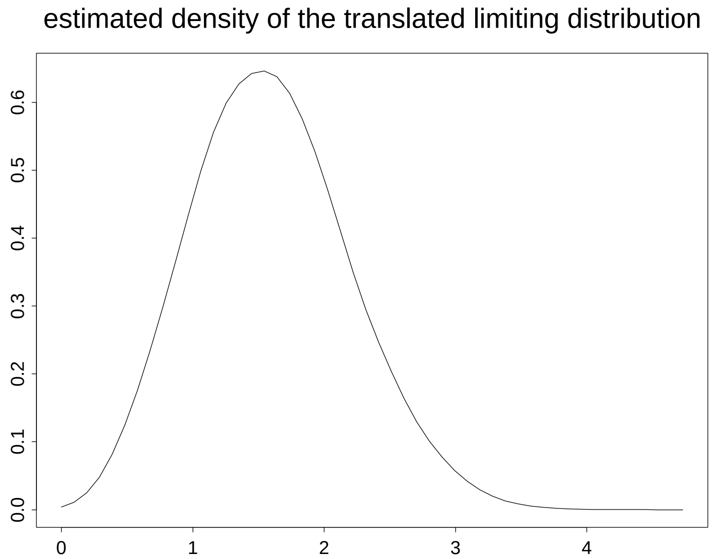
ξ being defined in (23); in particular $\xi^2 < 1$. Thus $(a_n)_{n\in\mathbb{N}}$ is bounded. Denote $a := \limsup_{n\to\infty} a_n$. Now we can conclude as in Rösler (1991). For a given $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ and $\xi^+ < 1$with $a_n \leq a + \epsilon$ and $\xi^2 + o(1) \leq \xi^+ < 1$ for all $n \geq n_0$. Then from (35) it follows that

$$\begin{aligned}
a_n &\leq 4\sum_{i=0}^{n_0-1} \mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}=i\}}\mathbf{1}_{\{Y<U\}}\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] a_i \\
&\quad + 4\sum_{i=n_0}^{n-1} \mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}=i\}}\mathbf{1}_{\{Y<U\}}\left(\frac{i}{n}\right)^{2\alpha-2} + o(1)\right] (a+\epsilon) + o(1) \\
&\leq \xi^+(a+\epsilon) + o(1). \qquad (37)
\end{aligned}$$

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

l_2-convergence implies weak convergence and convergence of the second moments. For this reason the constant β in (4) can be rederived from the limiting equation (30). We will give this argument in detail in the general d-dimensional case with $1 \leq s \leq d-1$ components specified in section 3. In this case the first order asymptotic for the variance has been unknown so far.

We cannot solve the limiting equation (30) explicitly. From a simulation we get an estimate for the density of the translated limiting distribution in (30).



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

3 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

$$\sum_{i=1}^{d} 2^{d-i} \mathbf{1}_{\{u_i \leq s_i\}}, \qquad u = (u_i), s = (s_i) \tag{38}$$

if s is a point in this subquadrant. A key p is inserted in the k–th subtree if it belongs to the k–th subquadrant. For the binary representation of $0 \leq$

$k \leq 2^d - 1$

$$k = \sum_{i=1}^{d} a_i 2^{d-i}, \qquad a_i = a_i(k) \in \{0, 1\} \tag{39}$$

let

$$E(k) := \{i \in \{1 \ldots, d\} \,|\, a_i(k) = 1\}, \tag{40}$$

$$N(k) := \{i \in \{1 \ldots, d\} \,|\, a_i(k) = 0\}. \tag{41}$$

Then equivalently, p is inserted in the k–th 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

$$\langle u \rangle_k := \prod_{i \in N(k)} u_i \prod_{i \in E(k)} (1 - u_i). \tag{42}$$

$\langle u \rangle := (\langle u \rangle_0, \ldots, \langle u \rangle_{2^d - 1})$ 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:

$$\mathbb{P}^{I^{(n)}|U=u} = M(n-1, \langle u \rangle), \tag{43}$$

where U, the first key to be inserted, is uniformly distributed on $[0,1]^d$. As in the 2-dimensional case convergence in probability of $I^{(n)}/n$ follows

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle U \rangle = (\langle U \rangle_0 \ldots, \langle U \rangle_{2^d - 1}). \tag{44}$$

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 $1 \leq s \leq d-1$ components specified we define for $j_1, \ldots, j_s \in \{0, 1\}$

$$\mathbf{1}_{j_1,\ldots,j_s}(U, Y) := \prod_{\substack{1 \leq i \leq s \\ j_i = 0}} \mathbf{1}_{\{Y_i < U_i\}} \prod_{\substack{1 \leq i \leq s \\ j_i = 1}} \mathbf{1}_{\{Y_i \geq U_i\}}. \tag{45}$$

Then analogously to (12)

$$C_n \stackrel{\mathcal{D}}{=} \sum_{j_1,\dots,j_s=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) \sum_{j_{s+1},\dots,j_d=0,1} C^{(j)}_{I^{(n)}_j} + 1. \tag{46}$$

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

$$j = \sum_{i=1}^{d} j_i 2^{d-i}. \tag{47}$$

In (46) $Y = (Y_1,\dots,Y_s), U = (U_1,\dots,U_d), (C_i^{(0)}),\dots,(C_i^{2^d-1})$ 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

$$\mathbb{E}C_n \sim \gamma_{s,d} n^{\alpha-1} \tag{48}$$

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*

$$\alpha^{d-s}(\alpha+1)^s = 2^d. \tag{49}$$

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

$$\mathrm{Var}(C_n) \sim \beta_{s,d} n^{2\alpha-2} \tag{50}$$

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

$$X_n := \frac{C_n - \mathbb{E}C_n}{n^{\alpha-1}} \tag{51}$$

with α given by (49) satisfies the modified recursion

$$\begin{aligned} X_n \stackrel{\mathcal{D}}{=} & \sum_{j_1,\dots,j_s=0,1} \mathbf{1}_{j_1,\dots,j_s}(U,Y) \\ & \times \sum_{j_{s+1},\dots,j_d=0,1} \left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} (X^{(j)}_{I^{(n)}_j} + \gamma_{s,d}) - \gamma_{s,d} + o(1). \end{aligned} \tag{52}$$

We define

$$U_{j_1,\ldots,j_d} := \prod_{\substack{1\le i\le d\\ j_i=0}} U_i \prod_{\substack{1\le i\le d\\ j_i=1}} (1-U_i) \tag{53}$$

for $j_1,\ldots,j_d \in \{0,1\}$. By the convergence of the coefficients in (44) it seems reasonable that a distributional limit X of (X_n) is a solution of the limiting equation

$$X \stackrel{\mathcal{D}}{=} \sum_{j_1,\ldots,j_s=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y) \sum_{j_{s+1},\ldots,j_d=0,1} U^{\alpha-1}_{j_1,\ldots,j_d}(X^{(j)}+\gamma_{s,d}) - \gamma_{s,d}. \tag{54}$$

Therefore, we define the *limiting operator*

$$T : M^1(I\!R^1,\mathcal{B}^1) \to M^1(I\!R^1,\mathcal{B}^1) \tag{55}$$

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

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

Lemma 3.1 *The limiting operator $T : M_{0,2} \to M_{0,2}$ is a contraction w.r.t. l_2:*

$$l_2(T(\mu),T(\nu)) \le \xi\, l_2(\mu,\nu) \quad \textit{for all} \quad \mu,\nu \in M_{0,2}, \tag{56}$$

$$\xi = \frac{1}{\sqrt{\alpha^s(\alpha-1/2)^{d-s}}} < 1. \tag{57}$$

Proof: Obviously $\mathrm{Var}(T(\mu)) < \infty$. Since the summands in (55) are identically distributed we derive

$$\begin{aligned} I\!ET(\mu) &= 2^d I\!E\left[\prod_{i=1}^s \mathbf{1}_{\{Y_i<U_i\}} \prod_{i=1}^d U_i^{\alpha-1}\right]\gamma_{s,d} - \gamma_{s,d} \\ &= 2^d I\!E\left[\mathbf{1}_{\{Y_1<U_1\}}U_1^{\alpha-1}\right]^s I\!E\left[U_1^{\alpha-1}\right]^{d-s}\gamma_{s,d} - \gamma_{s,d} \\ &= 2^d(\alpha+1)^{-s}\alpha^{-(d-s)}\gamma_{s,d} - \gamma_{s,d} \\ &= 2^d 2^{-d}\gamma_{s,d} - \gamma_{s,d} = 0, \end{aligned} \tag{58}$$

where the indicial equation (49) 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. $W^{(k)} \stackrel{\mathcal{D}}{=} \mu, Z^{(k)} \stackrel{\mathcal{D}}{=} \nu$ and $l_2^2(\mu,\nu) = \mathbb{E}(W^{(k)} - Z^{(k)})^2$ for $k = 0, \ldots, 2^d - 1$. Then using the independence properties and $\mathbb{E}W^{(k)} = \mathbb{E}Z^{(k)} = 0$ we conclude similarly to (24)

$$\begin{aligned}
& l_2^2(T(\mu), T(\nu)) \\
\leq\ & \mathbb{E}\left[\sum_{j_1,\ldots,j_s=0,1} \mathbf{1}_{j_1,\ldots,j_s}(Y,U) \sum_{j_{s+1},\ldots,j_d=0,1} U_{j_1,\ldots,j_d}^{2\alpha-2}(Z^{(j)} - W^{(j)})^2\right] \\
=\ & 2^d \mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i < U_i\}} U_i^{2\alpha-2} \prod_{i=s+1}^{d} U_i^{2\alpha-2}\right] l_2^2(\mu,\nu) \\
=\ & 2^d (2\alpha)^{-s} (2\alpha - 1)^{-(d-s)} l_2^2(\mu,\nu) \\
=\ & \frac{1}{\alpha^s(\alpha - 1/2)^{d-s}} l_2^2(\mu,\nu).
\end{aligned} \tag{59}$$

This implies assertion (56). Since $\alpha \in (1,2)$ we have $2\alpha - 1 > \alpha$ and $2\alpha > \alpha + 1$ which together with (49) yields

$$\begin{aligned}
\xi\ &=\ \left(\alpha^s(\alpha - 1/2)^{d-s}\right)^{-1/2} = \left(\frac{2^d}{(2\alpha)^s(2\alpha-1)^{d-s}}\right)^{1/2} \\
&<\ \frac{2^d}{(\alpha+1)^s \alpha^{d-s}} = 1.
\end{aligned} \tag{60}$$

■

As in the 2-dimensional case a simplification of the limiting operator T is possible (cf. (27),(28)). Denote

$$\bar{U}_{j_{s+1},\ldots,j_d} := \prod_{\substack{s+1 \leq i \leq d \\ j_i = 0}} U_i \prod_{\substack{s+1 \leq i \leq d \\ j_i = 1}} (1 - U_i). \tag{61}$$

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

$$T(\mu) \stackrel{\mathcal{D}}{=} \prod_{i=1}^{s} U_i^{\frac{\alpha-1}{2}} \left(\sum_{j_{s+1},\ldots,j_d=0,1} \bar{U}_{j_{s+1},\ldots,j_d}^{\alpha-1}(X^{(j_{s+1},\ldots,j_d)} + \gamma_{s,d})\right) - \gamma_{s,d}, \tag{62}$$

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 $j_{s+1},\dots,j_d = 0,1$. With an additional translation this again can be written more concisely. X is a fixed point of T in $M_{0,2}$ if and only if $\tilde{X} := X + \gamma_{s,d}$ is a fixed point of the operator $\tilde{T}$ on $M_{\gamma_{s,d},2}$ given by

$$\tilde{T}(\mu) :\stackrel{\mathcal{D}}{=} \prod_{i=1}^{s} U_i^{\frac{\alpha-1}{2}} \sum_{j_{s+1},\dots,j_d=0,1} \bar{U}_{j_{s+1},\dots,j_d}^{\alpha-1} \tilde{X}^{(j_{s+1},\dots,j_d)}. \tag{63}$$

In (63) again $\{U, \tilde{X}^{(j_{s+1},\dots,j_d)} : j_{s+1},\dots,j_d = 0,1\}$ is independent, U uniformly distributed on $[0,1]^d$ and $\tilde{X}^{(j_{s+1},\dots,j_d)} \stackrel{\mathcal{D}}{=} \mu \in M_{\gamma_{s,d},2}$.

Theorem 3.2 (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 (55), i.e.*

$$l_2(X_n, X) \to 0.$$

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

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

given in (63).

Proof: Using (44),(52) and (56) the proof of the 2-dimensional case can be extended to the multidimensional case. With $a_n := l_2^2(X_n, X)$ analogously to (31)–(35) the recursion

$$a_n \le 2^d \sum_{j=0}^{n-1} \mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i < U_i\}} \mathbf{1}_{\{I_1^{(n)} = j\}} (j/n)^{2\alpha-2} + o(1)\right] a_i + o(1) \tag{64}$$

can be derived. Then as in (36)

$$\begin{aligned} a_n &\le 2^d \left(\mathbb{E}\left[\prod_{i=1}^{s} \mathbf{1}_{\{Y_i < U_i\}} U_i^{2\alpha-2} \prod_{i=s+1}^{d} U_i^{2\alpha-2}\right] + o(1)\right) \sup_{1 \le i \le n-1} a_i + o(1) \\ &= (\xi^2 + o(1)) \sup_{1 \le i \le n-1} a_i + o(1) \end{aligned} \tag{65}$$

with ξ given in (57). This implies that $(a_n)_{n\in \mathbb{N}}$ is bounded. The convergence then follows as in (37). ■

Convergence in the l_2-metric implies convergence of the second moments (cf. (21)). 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 (63).

Corollary 3.3 (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*

$$\beta_{s,d} := \left[\frac{((2\alpha-1)B(\alpha,\alpha)+1)^{d-s}-1}{\alpha^s(\alpha-1/2)^{d-s}-1} - 1\right]\gamma_{s,d}. \tag{66}$$

The variance of the number of traversed nodes C_n satisfies

$$Var(C_n) \sim \beta_{s,d} n^{2\alpha-2}. \tag{67}$$

The constants α and $\gamma_{s,d}$ are given by (48)-(49) and $B(\cdot,\cdot)$ denotes the Eulerian beta integral.

Proof: Observe that

$$\mathrm{Var}(X) = \mathrm{Var}(\tilde{X}) = \mathbb{E}\tilde{X}^2 - \gamma_{s,d}^2 \tag{68}$$

where $\tilde{X}$ is the fixed point of the operator in (63). From (63) we obtain

$$\begin{aligned}
\mathbb{E}\tilde{X}^2 &= \mathbb{E}\left[\prod_{i=1}^s U_i^{\alpha-1} \sum_{\substack{j_{s+1},\dots,j_d=0,1\\ k_{s+1},\dots,k_d=0,1}} \bar{U}_{j_{s+1},\dots,j_d}^{\alpha-1}\bar{U}_{k_{s+1},\dots,k_d}^{\alpha-1}\tilde{X}^{(j_{s+1},\dots,j_d)}\tilde{X}^{(k_{s+1},\dots,k_d)}\right]\\
&= \alpha^{-s}\left[\sum_{\forall\, i:\, j_i=k_i} \mathbb{E}\bar{U}_{j_{s+1},\dots,j_d}^{2\alpha-2}\mathbb{E}\tilde{X}^2\right.\\
&\qquad \left. + \sum_{\exists\, i:\, j_i\neq k_i} \mathbb{E}\left[\bar{U}_{j_{s+1},\dots,j_d}^{\alpha-1}\bar{U}_{k_{s+1},\dots,k_d}^{\alpha-1}\right](\mathbb{E}\tilde{X})^2\right].
\end{aligned} \tag{69}$$

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

$$\mathbb{E}\bar{U}^{2\alpha-2}_{j_{s+1},\dots,j_d} = (2\alpha-1)^{-(d-s)} \tag{70}$$

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

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

$$\begin{aligned} \mathbb{E}\left[\bar{U}^{\alpha-1}_{j_{s+1},\dots,j_d}\bar{U}^{\alpha-1}_{k_{s+1},\dots,k_d}\right] &= \left(\mathbb{E}U_1^{2\alpha-2}\right)^{d-s-h}\left(\mathbb{E}(U_1(1-U_1))^{\alpha-1}\right)^h \\ &= (2\alpha-1)^{-(d-s-h)}(B(\alpha,\alpha))^h. \end{aligned} \tag{71}$$

With (71) in (69) we derive

$$\begin{aligned} \mathbb{E}\tilde{X}^2 &= \alpha^{-s}\Big[2^{d-s}(2\alpha-1)^{-(d-s)}\mathbb{E}\tilde{X}^2 \\ &\quad + \sum_{h=1}^{d-s} 2^{d-s}\binom{d-s}{h}(2\alpha-1)^{-(d-s-h)}(B(\alpha,\alpha))^h\gamma^2_{s,d}\Big]. \end{aligned} \tag{72}$$

Using the binomial formula it follows

$$\begin{aligned} &\mathbb{E}\tilde{X}^2\left(1-\alpha^{-s}\left(\frac{2}{2\alpha-1}\right)^{d-s}\right) \\ &= \alpha^{-s}2^{d-s}\left[\left(B(\alpha,\alpha)+\frac{1}{2\alpha-1}\right)^{d-s} - \left(\frac{1}{2\alpha-1}\right)^{d-s}\right]\gamma^2_{s,d}. \end{aligned} \tag{73}$$

A simplification leads to

$$\mathbb{E}\tilde{X}^2 = \frac{((2\alpha-1)B(\alpha,\alpha)+1)^{d-s}-1}{\alpha^s(\alpha-1/2)^{d-s}-1}\gamma^2_{s,d}. \tag{74}$$

Together with (68) this implies the first assertion.
By convergence of the second moments of X_n we finally conclude

$$\begin{aligned} \operatorname{Var}(C_n) &= \operatorname{Var}(n^{\alpha-1}X_n) = \operatorname{Var}(X_n)n^{2\alpha-2} = (\operatorname{Var}(X)+o(1))n^{2\alpha-2} \\ &\sim \beta_{s,d}n^{2\alpha-2}. \end{aligned} \tag{75}$$

■

The weak convergence in Theorem 3.2 also leads to results on exponential moments using the tools of Rösler (1991, 1992):

Theorem 3.4 (Convergence of Laplace transforms) *The limit X of the normalized number X_n of nodes traversed during a partial match query in a random d-dimensional quadtree with $1 \leq s \leq d-1$ components specified has a finite Laplace transform in some neighborhood of* 0*,*

$$\mathbb{E}\exp(\lambda X) < \infty \quad \textit{for all} \quad \lambda \in (-\lambda_0, \lambda_0). \tag{76}$$

For

$$0 < \frac{s}{d} < \frac{\ln(4/3)}{\ln(5/3)} = 0.563\ldots \tag{77}$$

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

$$\mathbb{E}\exp(\lambda X_n) \longrightarrow \mathbb{E}\exp(\lambda X), \quad \lambda \in \mathbb{R}. \tag{78}$$

Proof: Observe that the recursions for X_n and X given in (52) and (54) can be written in the form

$$X_n \stackrel{\mathcal{D}}{=} \sum_{j_1,\ldots,j_d=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y)\left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} X^{(j)}_{I_j^{(n)}} + C_n(U,Y,I^{(n)}) \tag{79}$$

and

$$X \stackrel{\mathcal{D}}{=} \sum_{j_1,\ldots,j_d=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y) U^{\alpha-1}_{j_1,\ldots,j_d} X^{(j)} + C(U,Y) \tag{80}$$

with

$$C_n(U,Y,I^{(n)}) = \sum_{j_1,\ldots,j_d=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y)\left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1} \gamma_{s,d} - \gamma_{s,d} + o(1) \tag{81}$$

and

$$C(U,Y) = \sum_{j_1,\ldots,j_d=0,1} \mathbf{1}_{j_1,\ldots,j_s}(U,Y) U^{\alpha-1}_{j_1,\ldots,j_d}\gamma_{s,d} - \gamma_{s,d}. \tag{82}$$

The distributions and (in-)dependencies are as in (52) and (54). The recursion (80) satsfies the conditions of Theorem 6 in Rösler (1992) with

$$T_j = \mathbf{1}_{j_1,\ldots,j_s}(U,Y)\left(\frac{I_j^{(n)}}{n}\right)^{\alpha-1}. \tag{83}$$

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

For the second assertion note that

$$\mathbb{E} C_n(U, Y, I^{(n)}) = 0 \quad \text{for all } n \in \mathbb{N} \tag{84}$$

since the X_n and $X^{(j)}_{I_j^{(n)}}$ in (79) are centered. Define

$$V_n := \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U, Y) \left(\frac{I_j^{(n)}}{n} \right)^{2\alpha-2} - 1. \tag{85}$$

It is $\sum_{j=0}^{2^d-1} I_j^{(n)} = n - 1$. The condition (77) and the indicial equation (49) imply $\alpha \geq 3/2$. Thus

$$V_n < 0 \quad \text{for all} \quad n \in \mathbb{N}. \tag{86}$$

The convergence of the coefficients in (44) implies

$$\begin{aligned} \mathbb{E} V_n \longrightarrow & \quad \mathbb{E} \sum_{j_1,\dots,j_d=0,1} \mathbf{1}_{j_1,\dots,j_s}(U, Y) U^{2\alpha-2}_{j_1,\dots,j_d} - 1 \\ = & \quad \xi - 1 < 0 \end{aligned} \tag{87}$$

with ξ given in (57). This yields

$$\sup_{n \in \mathbb{N}} \mathbb{E} V_n < 0. \tag{88}$$

From the representation (81) of $C_n(U, Y, I^{(n)})$ it is obvious that

$$\sup_{n \in \mathbb{N}} \|C_n\|_\infty < \infty. \tag{89}$$

The properties (84),(86),(88) and (89) are sufficient to obtain

$$\mathbb{E} \exp(\lambda X_n) \longrightarrow \mathbb{E} \exp(\lambda X), \quad \lambda \in \mathbb{R}. \tag{90}$$

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

In particular Theorem 3.4 implies exponential tails and the existence of all moments of the limiting distributions. Under condition (77) 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:

$$\mathbb{P}(C_n \geq a_n) \leq c_\lambda \exp\left(-\lambda \frac{a_n}{n^{\alpha-1}}\right). \tag{91}$$

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 (63)) are supported by $[0, \infty)$. The densities are positive almost everywhere on $[0, \infty)$ (cf. Theorem 2.4 in Tan and Hadjicostas (1995)).

References

Dobrow, R. P. and J. A. Fill (1999). Total path length for random recursive trees. To appear in *Combinatorics, Probability and Computing*.

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 (1991). The analysis of multidimensional searching in quad-trees. In *Proceedings of the Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 100–109.

Flajolet, P., G. Gonnet, C. Puech, and J. Robson (1993). Analytic variations on quadtrees. *Algorithmica 10*, 473–500.

Flajolet, P. and C. Puech (1986). Partial match retrieval of multidimensional data. *Journal of the ACM 33*, 371–407.

Knuth, D. (1998). *The art of computer programming: Sorting and searching* (second ed.), Volume 3. John Wiley.

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.

Martínez, C., A. Panholzer, and H. Prodinger (1998). Partial match queries in relaxed multidimensional search trees. Preprint.

Neininger, R. (1999). Asymptotic distributions for partial match queries in k-d trees. Preprint.

Neininger, R. and L. Rüschendorf (1999). On the internal path length of d-dimensional quad trees. *Random Structures and Algorithms 15*, 25–41.

Rachev, S. (1991). *Probability metrics and the stability of stochastic models.* John Wiley.

Rachev, S. and L. Rüschendorf (1995). Probability metrics and recursive algorithms. *Advances in Applied Probability 27*, 770–799.

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. (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.

Tan, K. H. and P. Hadjicostas (1995). Some properties of a limiting distribution in quicksort. *Statistics & Probability Letters 25*, 87–94.

Institut für Mathematische Stochastik
Universität Freiburg
Eckerstr. 1
79104 Freiburg
Germany
E-mail: {rn, ruschen}@stochastik.uni-freiburg.de