A limit process for partial match queries in random quadtrees and 2-d trees*

N. Broutin† R. Neininger‡ H. Sulzbach†

November 29, 2012

Abstract

We consider the problem of recovering items matching a partially specified pattern in multidimensional trees (quadtrees and k-d trees). We assume the traditional model where the data consist of independent and uniform points in the unit square. For this model, in a structure on n points, it is known that the number of nodes $C_n(\xi)$ to visit in order to report the items matching a random query ξ, independent and uniformly distributed on $[0, 1]$, satisfies $\mathbf{E}[C_n(\xi)] \sim \kappa n^\beta$, where κ and β are explicit constants. We develop an approach based on the analysis of the cost $C_n(s)$ of any fixed query $s \in [0, 1]$, and give precise estimates for the variance and limit distribution of the cost $C_n(x)$. Our results permit to describe a limit process for the costs $C_n(x)$ as x varies in $[0, 1]$; one of the consequences is that $\mathbf{E}[\max_{x\in[0,1]} C_n(x)] \sim \gamma n^\beta$; this settles a question of Devroye [Pers. Comm., 2000].

1 Introduction

Geometric databases arise in a number of contexts such as computer graphics, management of geographical data or statistical analysis. The aim consists in retrieving the data matching specified patterns efficiently. We are interested in tree-like data structures which permit such efficient searches. When the pattern specifies precisely all the data fields (we are looking for an *exact match*), the query can generally be answered in time logarithmic in the size of the database, and many precise analyses are available in this case, see, e.g., [16, 18, 20, 24, 25]. When the pattern only constrains some of the data fields (we are looking for a *partial match*), the searches must explore multiple branches of the data structure to report the matching data, and the cost usually becomes polynomial.

The first investigations about partial match queries by Rivest [34] were based on digital data structures (based on bit-comparisons). In a comparison-based setting, where the data may be compared directly at unit cost, a few general purpose data structures generalizing binary search trees permit to answer partial match queries, namely the quadtree [15], the k-d tree [1] and the relaxed k-d tree [11]. Besides the interest that one might have in partial match for its own sake, there are various reasons that justify the precise quantification of the cost of such general search queries in comparison-based data structures. First, these multidimensional trees are data structures of choice for applications that range from collision detection in motion planning to mesh generation [22, 41]. Furthermore, the cost of partial match queries also appears in (hence influences) the complexity of a number of other geometrical search questions such as range search [10] or rank selection [12]. For general references on multidimensional data structures and more details about their various applications, see the series of monographs by Samet [38, 39, 40].

In this paper, we provide refined analyses of the costs of partial match queries in some of the most important two dimensional data structures. We mostly focus on quadtrees. We extend our results to the case of k-d trees in Section 7. Similar results also hold for relaxed k-d trees of Duch et al. [11]. However, even stating them carefully would require much space without shedding anymore light on the phenomena, and we leave the straightforward modifications to the interested reader.

*Supported by MAEE and MESR Procope grant 23133PG (MAEE and MESR) and DAAD Procope project 50085686.
†Projet Algorithms, Inria Rocquencourt, 78153 Le Chesnay – France. Email: nicolas.broutin@inria.fr
‡J.W. Goethe-Universität Institut für Mathematik 60054 Frankfurt a.M. – Germany. Email: {neiningr,sulzbach}@math.uni-frankfurt.de

QUAD TREES AND MULTIDIMENSIONAL SEARCH. The quadtree [15] allows to manage multidimensional data by extending the divide-and-conquer approach of the binary search tree. Consider the point sequence $p_1, p_2, \ldots, p_n \in [0,1]^2$. As we build the tree, regions of the unit square are associated to the nodes where the points are stored. Initially, the root is associated with the region $[0,1]^2$ and the data structure is empty. The first point p_1 is stored at the root, and divides the unit square into four regions $Q_1, \ldots, Q_4$. Each region is assigned to a child of the root. More generally, when i points have already been inserted, we have a set of $1+3i$ (lower-level) regions that cover the unit square. The point p_{i+1} is stored in the node (say u) that corresponds to the region it falls in, divides it into four new regions that are assigned to the children of u. See Figure 1.

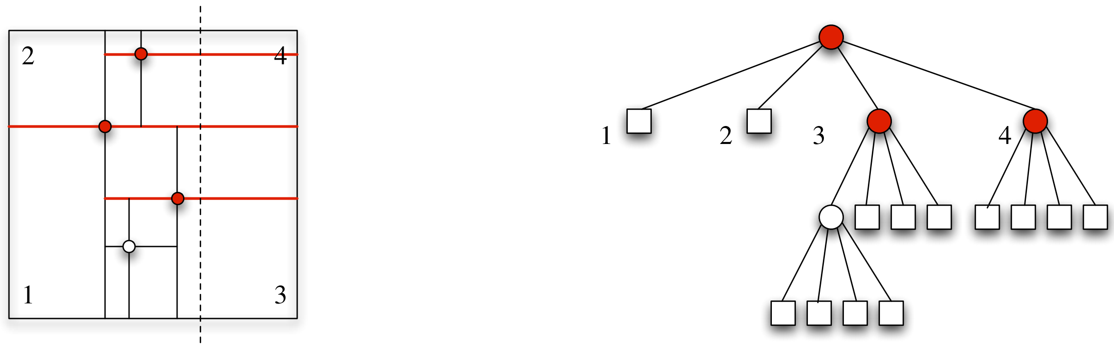

Figure 1: An example of a (point) quadtree: on the left the partition of the unit square induced by the tree data structure on the right (the children are ordered according to the numbering of the regions on the left). Answering the partial match query materialized by the dashed line on the left requires to visit the points/nodes coloured in red. Note that each one of the visited nodes correspond to a horizontal line that is crossed by the query.

ANALYSIS OF PARTIAL MATCH RETRIEVAL. For the analysis, we will focus on the model of *random quadtrees*, where the data points are independent and uniformly distributed in the unit square. In the present case, the data are just points, and the problem of partial match retrieval consists in reporting all the data with one of the coordinates (say the first) being $s \in [0,1]$. It is a simple observation that the number of nodes of the tree visited when performing the search is precisely $C_n(s)$, the number of regions in the quadtree that intersect a vertical line at s. The first analysis of partial match in quadtrees is due to Flajolet et al. [19] (after the pioneering work of Flajolet and Puech [17] in the case of k-d trees). They studied the singularities of a differential system for the generating functions of partial match cost to prove that, for a random query ξ, being independent of the tree and uniformly distributed on $[0,1]$, one has $\mathbf{E}[C_n(\xi)] \sim \kappa\, n^\beta$ where

$$\kappa = \frac{\Gamma(2\beta+2)}{2\Gamma(\beta+1)^3}, \qquad \text{and} \qquad \beta = \frac{\sqrt{17}-3}{2}, \tag{1}$$

and $\Gamma(x)$ denotes the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. Flajolet et al. [19] actually proved a more precise version of this estimate which will be crucial for us:

$$\mathbf{E}[C_n(\xi)] = \kappa\, n^\beta - 1 + O(n^{\beta-1}). \tag{2}$$

(This may also be obtained from the explicit expression for $\mathbf{E}[C_n(\xi)]$ devised by Chern and Hwang [4].)

Our aim in this paper is to gain a refined understanding of the cost beyond the level of expectations. In order to quantify the order of typical deviations from the mean we study the order of the variance together with limit distributions. However, deriving higher moments turns out to be subtle. In particular, when the query line is random (like above) although the four subtrees at the root are independent given their sizes, the contributions of the two subtrees that *do hit* the query line are *dependent*. Indeed, the relative location of the query line inside these two subtrees is again uniform, but unfortunately it is same in both regions. Hence, one cannot easily setup recurrence relations and perform an asymptotic analysis exploiting independence. This issue has not yet been addressed appropriately, and there is currently no result on the variance or higher moments for $C_n(\xi)$.

Another issue lies in the definition of the cost measure itself: even if the data follow some distribution, should one assume that the query follows the same distribution? In other words, should we focus on $C_n(\xi)$? Maybe not. But then, what distribution should one use for the query line?

One possible approach to overcome both problems is to consider the query line to be fixed and to study $C_n(s)$ for $s \in [0,1]$. This raises another problem: even if s is fixed at the top level, as the search is performed, the relative location of the queries in the recursive calls varies from one node to another. Thus, in following this approach, one is led to consider the entire stochastic process $(C_n(s))_{s\in[0,1]}$; this is the method we use here.

Recently Curien and Joseph [6] obtained some results in this direction. They proved that for every fixed $s \in (0,1)$,

$$\mathbf{E}[C_n(s)] \sim K_1 \cdot h(s) n^\beta \qquad \text{with} \qquad K_1 = \frac{\Gamma(2\beta+2)\Gamma(\beta+2)}{2\Gamma(\beta+1)^3\Gamma(\beta/2+1)^2}, \tag{3}$$

where the function h defined below will play a central role in the entire study

$$h(s) := (s(1-s))^{\beta/2}. \tag{4}$$

On the other hand, Flajolet et al. [19, 20] prove that, along the edge one has $\mathbf{E}[C_n(0)] = \Theta(n^{\sqrt{2}-1})$, so that $\mathbf{E}[C_n(0)] = o(n^\beta)$ (see also [6]). The behavior about the x-coordinate U of the first data point certainly resembles that along the edge, so that one has $\mathbf{E}[C_n(U)] = o(n^\beta)$. This suggests that $C_n(s)$ should not be concentrated around its mean, and that $n^{-\beta}C_n(s)$ should converge to a non-degenerate random variable as $n \to \infty$. Below, we confirm this and prove a functional limit law for $(n^{-\beta}C_n(s))_{s\in[0,1]}$ and characterize the limit process. From this we obtain refined asymptotic information on the complexity of partial match queries in quadtrees.

2 Main results and implications

We denote by $\mathcal{D}[0,1]$ the space of càdlàg functions on $[0,1]$ and by $\|f\| := \sup_{t\in[0,1]} |f(t)|$ the uniform norm of $f \in \mathcal{D}[0,1]$. Our main contribution is to prove the following convergence result:

Theorem 1. *Let $C_n(s)$ be the cost of a partial match query at a fixed line s in a random quadtree. Then, there exists a random continuous function Z such that, as $n \to \infty$,*

$$\left(\frac{C_n(s)}{K_1 n^\beta}, s \in [0,1]\right) \xrightarrow{d} (Z(s), s \in [0,1]). \tag{5}$$

This convergence in distribution holds in $\mathcal{D}[0,1]$ equipped with the Skorokhod topology.

The limit process Z may be characterized as follows (see Figure 2 for a simulation):

Proposition 2. *The distribution of the random function Z in (5) is a fixed point of the following functional recursive distributional equation, as process in $s \in [0,1]$,*

$$\begin{aligned} Z(s) \stackrel{d}{=} &\mathbf{1}_{\{s<U\}}\left[(UV)^\beta Z^{(1)}\left(\frac{s}{U}\right) + (U(1-V))^\beta Z^{(2)}\left(\frac{s}{U}\right)\right] \\ &+ \mathbf{1}_{\{s\geq U\}}\left[((1-U)V)^\beta Z^{(3)}\left(\frac{s-U}{1-U}\right) + ((1-U)(1-V))^\beta Z^{(4)}\left(\frac{s-U}{1-U}\right)\right], \end{aligned} \tag{6}$$

where U and V are independent $[0,1]$-uniform random variables and $Z^{(i)}$, $i = 1,\dots,4$ are independent copies of the process Z, which are also independent of U and V. Furthermore, Z in (5) is the only continuous solution of (6) with $\mathbf{E}[\|Z\|^2] < \infty$ and $\mathbf{E}[Z(\xi)] = \Gamma(\beta/2+1)^2/\Gamma(\beta+2)$ where ξ is independent of Z and uniformly distributed on $[0,1]$.

The methods applied to prove Theorem 1 also guarantee convergence of the variance of the costs of partial match queries. The following theorem for uniform queries ξ is the direct extension of the pioneering work in [17, 19] for the cost of partial match queries at a uniform line ξ in random two-dimensional trees.

Theorem 3. *If ξ is uniformly distributed on $[0,1]$, independent of (C_n) and Z, then as $n \to \infty$,*

$$\frac{C_n(\xi)}{K_1 n^\beta} \to Z(\xi)$$

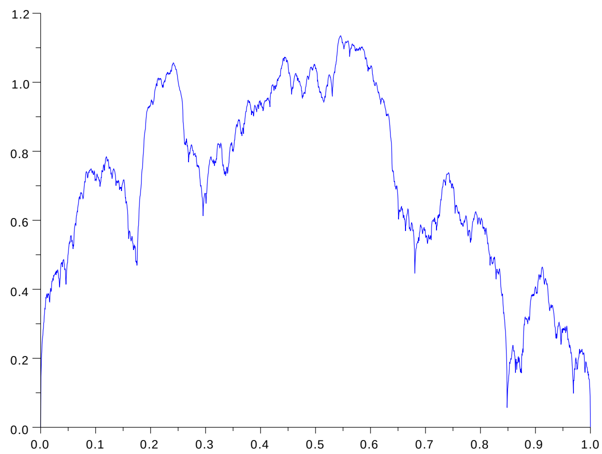

Figure 2: A simulation of the limit process Z.

in distribution. Moreover, $\mathbf{Var}\,(C_n(\xi)) \sim K_4 n^{2\beta}$ *where* $K_4 \approx 0.447363034$ *is given by, with* K_1 *in (3),*

$$K_4 := K_1^2 \cdot \mathbf{Var}(Z(\xi)) = K_1^2 K_3 = K_1^2 \left(\frac{2(2\beta+1)}{3(1-\beta)} \mathrm{B}(\beta+1,\beta+1)^2 - \mathrm{B}(\beta/2+1,\beta/2+1)^2 \right) \quad (7)$$

Here $\mathrm{B}(a,b) := \int_0^1 t^{a-1}(1-t)^{b-1}dt$ denotes the Eulerian integral for $a,b > -1$. In particular, Theorem 3 identifies the first-order asymptotics of $\mathbf{Var}(C_n(\xi))$ which is to be compared with studies that neglected the dependence between the contributions of the subtrees mentioned above [26, 27, 29]. A refined result about the variance $\mathbf{Var}(C_n(s))$ at a fixed location reads

$$\mathbf{Var}\,(C_n(s)) \sim K_1^2 \mathbf{Var}\,(Z(s))\, n^{2\beta},$$

where $s \in (0,1)$ and an explicit expression for $\mathbf{Var}\,(Z(s))$ is given by

$$\mathbf{Var}\,(Z(s)) = K_2 h^2(s) = \left[2B\,(\beta+1,\beta+1)\,\frac{2\beta+1}{3(1-\beta)} - 1 \right] h^2(s). \quad (8)$$

Another consequence of Theorem 1 concerns the order of the cost of the worst query $\sup_{s\in[0,1]} C_n(s)$.

Theorem 4. *Let* $S_n = \sup_{s\in[0,1]} C_n(s)$. *Then, as* $n \to \infty$,

$$\frac{S_n}{K_1 n^\beta} \to S := \sup_{s\in[0,1]} Z(s),$$

in distribution and with convergence of all moments. In particular, $\mathbf{E}[S] < \infty$, $\mathbf{Var}(S) < \infty$ *and we have*

$$\mathbf{E}[S_n] \sim K_1 n^\beta \mathbf{E}[S], \qquad \textit{and} \qquad \mathbf{Var}(S_n) \sim K_1^2 n^{2\beta} \mathbf{Var}(S).$$

Note that the sequence $n^{-\beta}\mathbf{E}[S_n]$ is bounded. In particular, $\mathbf{E}[S_n]$ has the same order of magnitude as the cost of a search query at any single location, and does not include any extra factor growing with n. Interestingly, the one-dimensional marginals of the limit process $(Z(s), s \in [0,1])$ are all the same up to a deterministic multiplicative constant given by the function h:

Theorem 5. *There exists a random variable* $\Psi \geq 0$ *such that for all* $s \in [0,1]$,

$$Z(s) \stackrel{d}{=} h(s) \cdot \Psi. \quad (9)$$

The distribution of Ψ is the unique solution of the fixed-point equation

$$\Psi \stackrel{d}{=} U^{\beta/2}V^{\beta}\Psi + U^{\beta/2}(1-V)^{\beta}\Psi \tag{10}$$

with $\mathbf{E}[\Psi] = 1$ and $\mathbf{E}[\Psi^2] < \infty$ where Ψ' is an independent copy of Ψ and (Ψ, Ψ') is independent of (U, V).

Convergence of all moments of the supremum $n^{-\beta}S_n$ in Theorem 4 implies uniform integrability of any moment of the process $n^{-\beta}C_n$, hence the following result about convergence of all moments.

Corollary 6. *For all $s \in [0, 1]$, we have*

$$\mathbf{E}\left[\left(\frac{C_n(s)}{K_1 n^{\beta}}\right)^m\right] \to \mathbf{E}\left[Z(s)^m\right] = c_m h(s)^m,$$

for all $m \in \mathbb{N}$ as $n \to \infty$ where c_m is given by

$$c_m = \frac{\beta m + 1}{(m-1)\left(m+1-\frac{3}{2}\beta m\right)} \sum_{\ell=1}^{m-1} \binom{m}{\ell} \mathrm{B}(\beta\ell+1, \beta(m-\ell)+1) c_\ell c_{m-\ell}, \tag{11}$$

for $m \geq 2$ where $c_1 = 1$. An analogous result holds true for $\mathbf{E}[C_n(\xi)]$ where ξ is uniform on $[0, 1]$ and independent of $(C_n)_{n\geq 0}$ and Z, and for moments involving queries at multiple locations.

Plan of the paper. Our approach requires to work with the process $(C_n(s) : s \in [0, 1])$ and is based on the recursive decomposition of the tree at the root. This yields a recursive distributional recurrence for $(C_n(s) : s \in [0, 1])$ to which we apply a functional version of the contraction method. In Section 3, we give an overview of this underlying methodology. In particular, we discuss the novel results of Neininger and Sulzbach [32] about the contraction method in function spaces which we will apply. Sections 4 and 5 are dedicated to the proofs of two of the main ingredients required to apply the results of [32], the existence of a continuous solution of the limit recursive equation, and the uniform convergence of the rescaled first moment $n^{-\beta}\mathbf{E}[C_n(s)]$ at an appropriate rate. In Section 6, we identify the variance and the supremum of the limit process Z, and deduce the large n asymptotics for $C_n(s)$ in Theorems 3 and 4. Finally, we prove analogous results for the cases of 2-d trees in Section 7. Our results on quadtrees have been announced in the extended abstract [3].

3 Contraction method in function spaces

3.1 Overview of the method

The aim of this section is to give an overview of the method we employ to prove Theorem 1. It is based on a contraction argument in a certain space of probability distributions. In the context of the analysis of algorithms, the method was first employed by Rösler [35] who proved convergence in distribution for the rescaled total cost of the randomized version of quicksort. The method was then further developed by Rösler [36], Rachev and Rüschendorf [33], and later on in [9, 13, 28, 30, 31, 37] and has permitted numerous analyses in distribution for random discrete structures.

So far, the method has mostly been used to analyze random variables taking real values, though a few applications on function spaces have been made, see [9, 13, 21]. Here we are interested in the function space $\mathcal{D}[0, 1]$ endowed with the Skorokhod topology [see, e.g., 2], but the main idea persists: (1) devise a recursive equation for the quantity of interest (here the process$(C_n(s), s \in [0, 1])$), and (2) based on a properly rescaled version of the quantity deduce a limit equation, i.e., a recursive distributional equation that the limit may satisfy; (3) if the map of distributions associated to the limit equation is a contraction in a certain metric space, then a fixed point is unique and may be obtained by iteration. The contraction may also be exploited to obtain weak convergence to the fixed point. We now move on to the first step of this program.

Write $I_1^{(n)}, \ldots, I_4^{(n)}$ for the number of points falling in the four regions created by the point stored at the root. Then, given the coordinates of the first data point (U, V), we have, cf. Figure 1,

$$(I_1^{(n)}, \ldots, I_4^{(n)}) \stackrel{d}{=} \mathrm{Mult}(n-1; UV, U(1-V), (1-U)(1-V), (1-U)V). \tag{12}$$

Observe that, for the cost inside a subregion, what matters is the location of the query line *relative* to the region. Thus a decomposition at the root yields the following recursive relation, for any $n \geq 1$,

$$C_n(s) \stackrel{d}{=} 1 + \mathbf{1}_{\{s<U\}} \left[C^{(1)}_{I_1^{(n)}}\Big(\frac{s}{U}\Big) + C^{(2)}_{I_2^{(n)}}\Big(\frac{s}{U}\Big) \right] + \mathbf{1}_{\{s\geq U\}} \left[C^{(3)}_{I_3^{(n)}}\Big(\frac{1-s}{1-U}\Big) + C^{(4)}_{I_4^{(n)}}\Big(\frac{1-s}{1-U}\Big) \right], \tag{13}$$

where $U, I_1^{(n)}, \dots, I_4^{(n)}$ are the quantities already introduced and $(C_k^{(1)}), \dots, (C_k^{(4)})$ are independent copies of the sequence $(C_k, k \geq 0)$, independent of $(U, V, I_1^{(n)}, \dots, I_4^{(n)})$. We stress that this equation does not only hold true pointwise for fixed s but also as càdlàg functions on the unit interval. The relation in (13) is the fundamental equation for us.

Letting $n \to \infty$ (formally) in (13) suggests that, if $n^{-\beta} C_n(s)$ does converge to a random variable $Z(s)$ in a sense to be made precise, then the distribution of the process $(Z(s), 0 \leq s \leq 1)$ should satisfy the following fixed point equation

$$\begin{aligned} Z(s) \stackrel{d}{=} &\mathbf{1}_{\{s<U\}} \left[(UV)^\beta Z^{(1)}\left(\frac{s}{U}\right) + (U(1-V))^\beta Z^{(2)}\left(\frac{s}{U}\right) \right] \\ &+ \mathbf{1}_{\{s\geq U\}} \left[((1-U)V)^\beta Z^{(3)}\left(\frac{s-U}{1-U}\right) + ((1-U)(1-V))^\beta Z^{(4)}\left(\frac{s-U}{1-U}\right) \right], \end{aligned} \tag{14}$$

where U and V are independent $[0,1]$-uniform random variables and $Z^{(i)}$, $i = 1, \dots, 4$ are independent copies of the process Z, which are also independent of U and V.

The last step leading to the fixed point equation (14) needs now to be made rigorous. It is at this point that the contraction method enters the game. The distribution of a solution to our fixed-point equation (14) lies in the set of probability measures on the Polish space $(\mathcal{D}[0,1], d)$, which is the set we have to endow with a suitable metric. Here, d denotes the Skorokhod metric [see, e.g., 2].

The recursive equation (13) is an example for the following, more general setting of random additive recurrences: Let (X_n) be $\mathcal{D}[0,1]$-valued random variables with

$$X_n \stackrel{d}{=} \sum_{r=1}^{K} A_r^{(n)}\Big(X^{(r)}_{I_r^{(n)}}\Big) + b^{(n)}, \quad n \geq 1, \tag{15}$$

where $(A_1^{(n)}, \dots, A_K^{(n)})$ are random continuous linear operators on $\mathcal{D}[0,1]$, $b^{(n)}$ is a $\mathcal{D}[0,1]$-valued random variable, $I_1^{(n)}, \dots, I_K^{(n)}$ are random integers between 0 and $n-1$ and the sequences of process $(X_n^{(1)}), \dots, (X_n^{(K)})$ are distributed like (X_n). Moreover $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I_1^{(n)}, \dots, I_K^{(n)}), (X_n^{(1)}), \dots, (X_n^{(K)})$ are independent.

At this point, one should comment on the term random continuous linear operator: As explained explicitly in [32], A is a random continuous linear operator on $\mathcal{D}[0,1]$, if it takes values in the set of endomorphisms on $\mathcal{D}[0,1]$ that are both continuous with respect to the supremum norm and to the Skorokhod metric. Moreover, for any $f \in \mathcal{D}[0,1]$ and $t \in [0,1]$, the quantity $Af(t)$ has to be a real-valued random variable, and the same is assumed for $\|A\|_{\mathrm{op}}$ (see below for the definition). Finally, we remember that convergence $d(f_n, f) \to 0$ in the Skorokhod metric means that there exists a sequence of monotonically increasing bijections (λ_n) on the unit interval such that $f_n(\lambda_n(t)) \to f(t)$ and $\lambda_n(t) \to t$ both uniformly in t as $n \to \infty$.

To establish Theorem 1 as a special case of this setting we use Proposition 7 below. Proposition 7 is part of the main convergence theorem in Neininger and Sulzbach [32]. We first state conditions needed to deal with the general recurrence (15); we will then justify that it can indeed be used in the case of cost of partial match queries. Consider the following assumptions, where, for a random variable X in $\mathcal{D}[0,1]$ we write $\|X\|_2 := \mathbf{E}[\|X\|^2]^{1/2}$, for a linear operator A we write $\|A\|_2 := \mathbf{E}[\|A\|_{\mathrm{op}}^2]^{1/2}$ with $\|A\|_{\mathrm{op}} := \sup_{\|x\|=1} \|A(x)\|$. Suppose (X_n) obeys (15) and

(A1) Convergence and contraction. We have $\|A_r^{(n)}\|_2, \|b^{(n)}\|_2 < \infty$ for all $r = 1, \dots, K$ and $n \geq 0$ and there exist random continuous linear operators $A_1, \dots, A_K$ on $\mathcal{D}[0,1]$ and a $\mathcal{D}[0,1]$-valued random variable b such that, for some positive sequence $R(n) \downarrow 0$, as $n \to \infty$,

$$\|b^{(n)} - b\|_2 + \sum_{r=1}^{K} \|A_r^{(n)} - A_r\|_2 = O(R(n)) \tag{16}$$

and for all $\ell \in \mathbb{N}$,

$$\mathbf{E}\left[\mathbf{1}_{\{I_r^{(n)} \in \{0,\dots,\ell\}\}} \|A_r^{(n)}\|_{\mathrm{op}}^2\right] \to 0$$

and

$$L^* = \limsup_{n\to\infty} \mathbf{E}\left[\sum_{r=1}^{K} \|A_r^{(n)}\|_{\mathrm{op}}^2 \frac{R(I_r^{(n)})}{R(n)}\right] < 1. \tag{17}$$

(A2) EXISTENCE AND EQUALITY OF MOMENTS. $\mathbf{E}[\|X_n\|^2] < \infty$ for all n and $\mathbf{E}[X_{n_1}(t)] = \mathbf{E}[X_{n_2}(t)]$ for all $n_1, n_2 \in \mathbb{N}_0, t \in [0,1]$.

(A3) EXISTENCE OF A CONTINUOUS SOLUTION. There exists a solution X of the fixed-point equation

$$X \stackrel{d}{=} \sum_{r=1}^{K} A_r(X^{(r)}) + b \tag{18}$$

with continuous paths, $\mathbf{E}[\|X\|^2] < \infty$ and $\mathbf{E}[X(t)] = \mathbf{E}[X_1(t)]$ for all $t \in [0,1]$. Again the random variables $(A_1,\dots,A_K,b), X^{(1)},\dots,X^{(K)}$ are independent and $X^{(1)},\dots,X^{(K)}$ are distributed like X.

(A4) PERTURBATION CONDITION. $X_n = W_n + h_n$ where $\|h_n - h\| \to 0$ with $h \in \mathcal{C}[0,1]$ and random variables W_n in $\mathcal{D}[0,1]$ such that there exists a sequence (r_n) with, as $n \to \infty$,

$$\mathbf{P}\left(W_n \notin \mathcal{D}_{r_n}[0,1]\right) \to 0.$$

Here, $\mathcal{D}_{r_n}[0,1] \subset \mathcal{D}[0,1]$ denotes the set of functions on the unit interval continuous at 1, for which there is a decomposition of $[0,1]$ into intervals of length as least r_n on which they are constant.

(A5) RATE OF CONVERGENCE. $R(n) = o\left(\log^{-2}(1/r_n)\right)$.

The contraction method presented here for the space $(\mathcal{D}[0,1], d)$ is based on the Zolotarev metric ζ_2, see [32]. We state the part of the main convergence theorem of Neininger and Sulzbach [32] that we will use. In the next section, we will prove our main result, Theorem 1, with the help of Proposition 7.

Proposition 7. *Let (X_n) fulfill (15). Provided that assumptions (A1)–(A3) are satisfied, the solution X of the fixed-point equation (18) is unique.*

i. For all $t \in [0,1]$, $X_n(t) \to X(t)$ in distribution, with convergence of the first two moments;

ii. If ξ is independent of (X_n), X and distributed on $[0,1]$ then $X_n(\xi) \to X(\xi)$ in distribution again with convergence of the first two moments.

iii. If also (A4) and (A5) hold, then $X_n \to X$ in distribution in $(\mathcal{D}[0,1], d)$.

Note that $X_n \to X$ in distribution in $(\mathcal{D}[0,1], d)$ with X having continuous sample paths implies that we can find versions of $(X_n), X$ on a suitable probability space such that $\|X_n - X\| \to 0$ almost surely. However, in general we do not have $X_n \to X$ in distribution in $\mathcal{D}[0,1]$ endowed with the uniform topology due to problems with measurability, see [2, Section 15] and [32, Section 2.2].

3.2 The functional limit theorem: Proof of Theorem 1

The aim of this section is to prove Theorem 1 with the help of Proposition 7 from Neininger and Sulzbach [32]. More precisely, in the following we prove all the conditions (A1)–(A5) except two which require much more work: the existence of a continuous solution (A3), and the uniform convergence of the mean in (A1) are treated separately in Sections 4 and 5, respectively.

Following the heuristics in the introduction we scale the additive recurrence (13) by n^β. Let $Q_0(t) := 0$ and

$$Q_n(t) = \frac{C_n(t)}{K_1 n^\beta}, \qquad n \geq 1.$$

The recursive distributional equation then rewrites in terms of Q_n as

$$(Q_n(t))_{t\in[0,1]} \stackrel{d}{=} \Bigg(\mathbf{1}_{\{t<U\}}\left[\left(\frac{I_1^{(n)}}{n}\right)^\beta Q^{(1)}_{I_1^{(n)}}\left(\frac{t}{U}\right)+\left(\frac{I_2^{(n)}}{n}\right)^\beta Q^{(2)}_{I_2^{(n)}}\left(\frac{t}{U}\right)\right] + \mathbf{1}_{\{t\ge U\}}\left[\left(\frac{I_3^{(n)}}{n}\right)^\beta Q^{(3)}_{I_3^{(n)}}\left(\frac{t-U}{1-U}\right)+\left(\frac{I_4^{(n)}}{n}\right)^\beta Q^{(4)}_{I_4^{(n)}}\left(\frac{t-U}{1-U}\right)\right] + \frac{1}{K_1 n^\beta}\Bigg)_{t\in[0,1]} \tag{19}$$

where $U, I_1^{(n)}, \dots, I_4^{(n)}$ are the quantities already introduced in Section 3.1 and (12) and $(Q_n^{(1)})_{n\ge0}, \dots, (Q_n^{(4)})_{n\ge0}$ are independent copies of $(Q_n)_{n\ge0}$, independent of $(U, V, I_1^{(n)}, \dots, I_4^{(n)})$. The convergence of the coefficients $(I_j^{(n)}/n)^\beta$ suggests that a limit of $Q_n(t)$ should satisfy the fixed-point equation (14).

THE RECURRENCE RELATION. Most details consist in setting the right form of the recurrence relation: for (A2) to be satisfied, we need to use a scaling that leads to an expectation which is independent of n. This is not the case for $Q_n(t)$. Denoting $\mu_n(t) = \mathbf{E}[C_n(t)]$, we are naturally led to consider $Y_0(t) := 0$ and

$$Y_n(t) = \frac{C_n(t)-\mu_n(t)}{K_1 n^\beta} = Q_n(t) - h(t) + O(n^{-\varepsilon}), \qquad n \ge 1.$$

where the error term is deterministic and uniform in $t \in [0,1]$. Hence it is sufficient to prove convergence of the sequence $(Y_n)_{n\ge1}$. The distributional recursion in terms of Y_n is

$$(Y_n(t))_{t\in[0,1]} \stackrel{d}{=} \Bigg(\mathbf{1}_{\{t<U\}}\left[\left(\frac{I_1^{(n)}}{n}\right)^\beta Y^{(1)}_{I_1^{(n)}}\left(\frac{t}{U}\right)+\left(\frac{I_2^{(n)}}{n}\right)^\beta Y^{(2)}_{I_2^{(n)}}\left(\frac{t}{U}\right)\right] + \mathbf{1}_{\{t\ge U\}}\left[\left(\frac{I_3^{(n)}}{n}\right)^\beta Y^{(3)}_{I_3^{(n)}}\left(\frac{t-U}{1-U}\right)+\left(\frac{I_4^{(n)}}{n}\right)^\beta Y^{(4)}_{I_4^{(n)}}\left(\frac{t-U}{1-U}\right)\right] + \mathbf{1}_{\{t<U\}}\left[\frac{\mu_{I_1^{(n)}}\left(\frac{t}{U}\right)+\mu_{I_2^{(n)}}\left(\frac{t}{U}\right)}{K_1 n^\beta}\right] + \mathbf{1}_{\{t\ge U\}}\left[\frac{\mu_{I_3^{(n)}}\left(\frac{t-U}{1-U}\right)+\mu_{I_4^{(n)}}\left(\frac{t-U}{1-U}\right)}{K_1 n^\beta}\right] + \frac{1-\mu_n(t)}{K_1 n^\beta}\Bigg)_{t\in[0,1]},$$

where $(Y_n^{(1)})_{n\ge0}, \dots, (Y_n^{(4)})_{n\ge0}$ are independent copies of $(Y_n)_{n\ge0}$ which are also independent of the vector $(U, V, I_1^{(n)}, \dots, I_4^{(n)})$. Therefore, any possible limit Y of Y_n should satisfy the following distributional fixed-point equation

$$(Y(t))_{t\in[0,1]} \stackrel{d}{=} \Bigg(\mathbf{1}_{\{t<U\}}\left[(UV)^\beta Y^{(1)}\left(\frac{t}{U}\right) + (U(1-V))^\beta Y^{(2)}\left(\frac{t}{U}\right)\right] + \mathbf{1}_{\{t\ge U\}}\left[((1-U)V)^\beta Y^{(3)}\left(\frac{t-U}{1-U}\right) + ((1-U)(1-V))^\beta Y^{(4)}\left(\frac{t-U}{1-U}\right)\right] + \mathbf{1}_{\{t\ge U\}} h\left(\frac{t-U}{1-U}\right)\left(((1-U)V)^\beta + ((1-U)(1-V))^\beta\right) - h(t) + \mathbf{1}_{\{t<U\}} h\left(\frac{t}{U}\right)\left((UV)^\beta + (U(1-V))^\beta\right)\Bigg)_{t\in[0,1]}. \tag{20}$$

Having Proposition 7 in mind, we define (random) operators $A_r^{(n)}$, $r = 1, 2, 3, 4$, by

$$A_r^{(n)}(f)(t) = \begin{cases} \mathbf{1}_{\{t<U\}}\left(\frac{I_r^{(n)}}{n}\right)^\beta f\left(\frac{t}{U}\right) & \text{if } r = 1, 2 \\ \mathbf{1}_{\{t\ge U\}}\left(\frac{I_r^{(n)}}{n}\right)^\beta f\left(\frac{t-U}{1-U}\right) & \text{if } r = 3, 4. \end{cases}$$

Furthermore let $b^{(n)}(t) = \sum_{r=1}^{4} b_r^{(n)}(t) + (1-\mu_n(t))/(K_1 n^\beta)$ with

$$b_r^{(n)}(t) = \begin{cases} \mathbf{1}_{\{t<U\}} \cdot \dfrac{\mu_{I_r^{(n)}}\left(\frac{t}{U}\right)}{K_1 n^\beta} & \text{if } r = 1,2 \\ \mathbf{1}_{\{t\geq U\}} \cdot \dfrac{\mu_{I_r^{(n)}}\left(\frac{t-U}{1-U}\right)}{K_1 n^\beta} & \text{if } r = 3,4. \end{cases}$$

Then the finite-n version of the recurrence relation for $(Y_n)_{n\geq 0}$ is precisely of the form (15).

We define similarly the coefficients of the limit recursive equation (20). We will then show that with these definitions, the assumptions (A1)–(A5) are satisfied (again, except the existence of a continuous limit solution and the uniform convergence for the mean treated in Section 4 and 5). The operators $A_1, \dots, A_4$ are defined by

$$A_1(f)(t) = \mathbf{1}_{\{t<U\}} (UV)^\beta f\left(\frac{t}{U}\right) \qquad A_2(f)(t) = \mathbf{1}_{\{t<U\}} (U(1-V))^\beta f\left(\frac{t}{U}\right)$$
$$A_3(f)(t) = \mathbf{1}_{\{t\geq U\}} ((1-U)V)^\beta f\left(\frac{t-U}{1-U}\right) \qquad A_4(f)(t) = \mathbf{1}_{\{t\geq U\}} ((1-U)(1-V))^\beta f\left(\frac{t}{U}\right)$$

and $b(t) = \sum_{r=1}^{4} b_r(t) - h(t)$ with

$$b_1(t) = \mathbf{1}_{\{t<U\}} (UV)^\beta h\left(\frac{t}{U}\right), \qquad b_2(t) = \mathbf{1}_{\{t<U\}} (U(1-V))^\beta h\left(\frac{t}{U}\right)$$
$$b_3(t) = \mathbf{1}_{\{t\geq U\}} ((1-U)V)^\beta h\left(\frac{t-U}{1-U}\right), \qquad b_4(t) = \mathbf{1}_{\{t\geq U\}} ((1-U)(1-V))^\beta h\left(\frac{t}{U}\right).$$

The operators $A_1, \dots, A_4, A_1^{(n)}, \dots, A_4^{(n)}$ are linear for each n. Moreover, they are bounded above by one which implies that they are norm-continuous. Their norm functions are real-valued random variables. In order to establish that they are indeed random continuous linear operators on $(\mathcal{D}[0,1], d)$ it remains to check that they are continuous with respect to the Skorokhod topology. To this end, it is sufficient to prove that

$$d(f_n, f) \to 0 \Rightarrow d\left(\mathbf{1}_{\{t<u\}} f_n\left(\frac{t}{u}\right), \mathbf{1}_{\{t<u\}} f\left(\frac{t}{u}\right)\right) \to 0$$

for any $u \in [0,1]$. This follows easily since $\|f_n(\lambda_n(t)) - f(t)\| \to 0$ with monotonically increasing bijections λ_n on the unit interval such that $\|\lambda_n(t) - t\| \to 0$ implies $\|\mathbf{1}_{\{\beta_n(t)<u\}} f_n(\beta_n(t)/u) - \mathbf{1}_{\{t<u\}} f(t/u)\| \to 0$ where $\beta_n(t) = u\,\lambda_n(t/u)$ for $t \leq u$ and $\beta_n(t) = t$ for $t > u$.

We are now ready to check that the assumptions (A1)–(A5) indeed hold, taking the results of Sections 4 and 5 for granted.

(A3) Existence of a continuous solution. In Section 4, we construct a continuous solution Z of the fixed-point equation (14) with $\mathbf{E}[\|Z\|^2] < \infty$ and $\mathbf{E}[Z(t)] = h(t) = (t(1-t))^{\beta/2}$. Hence the function $Y(t) = Z(t) - h(t)$ is a continuous solution of (20) with $\mathbf{E}[Y(t)] = 0$ and $\mathbf{E}[\|Y\|^2] < \infty$. A direct computation shows that $\mathbf{E}[\|A_r\|_{\text{op}}^2] = \mathbf{E}[(UV)^{2\beta}] = (2\beta+1)^{-2}$, for $r = 1, \dots, 4$. Observe that

$$L := \sum_{r=1}^{4} \mathbf{E}[\|A_r\|_{\text{op}}^2] = \frac{4}{(2\beta+1)^2} < 1.$$

In particular, Y is the unique solution of (20) with $\mathbf{E}\left[Y(t)\right] = 0$ and $\mathbf{E}[\|Y\|^2] < \infty$. Thus, Z is the unique solution of (6) with $\mathbf{E}[Z(t)] = h(t)$ and $\mathbf{E}[\|Z\|^2] < \infty$. By the arguments in [6, Section 5], the mean function of any process with càdlàg paths and finite moments satisfying (6) is a multiple of $h(s)$. Hence, we may replace the condition $\mathbf{E}\left[Z(t)\right] = h(t)$ by $\mathbf{E}\left[Z(\xi)\right] = \Gamma(\beta/2+1)^2/\Gamma(\beta+2)$ as formulated in Proposition 2.

(A2) Existence and equality of moments. The precise scaling we chose ensures that $\mathbf{E}[Y_n(t)] = 0$, for all $n \geq 1$ and $t \in [0,1]$. The second moments $\mathbf{E}[\|Y_n\|^2]$ are finite as the random variables $\|Y_n\|$ are bounded for every fixed n.

(A1) CONVERGENCE AND CONTRACTION. It suffices to focus on the terms

$$\|A_1^{(n)} - A_1\|_2 \qquad \text{and} \qquad \|b_1^{(n)} - b_1\|_2,$$

the remaining terms can obviously be treated in the same way. Establishing the convergence only boils down to verifying that a binomial random variable $\mathrm{Bin}(n,p)$ is properly approximated by np. Using the Chernoff–Hoeffding inequality for binomials [23], one easily verifies that for every $\alpha > 0$,

$$\mathbf{E}\left[\left|\frac{\mathrm{Bin}(n,p)}{n} - p\right|^{\alpha}\right] = O(n^{-\alpha/2}), \tag{21}$$

uniformly in $p \in [0,1]$. Thus, since $|x^\beta - y^\beta| \le |x-y|^\beta$ for any $x, y \in [0,1]$, we have

$$\|A_1^{(n)} - A_1\|_2 \le \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\beta} - (UV)^\beta \right\|_2 = O(n^{-1/2}). \tag{22}$$

By Proposition 12 we have $\mu_n(t) = K_1 h(t) n^\beta + O(n^{\beta-\varepsilon})$ uniformly in $t \in [0,1]$. Therefore

$$\|b_1^{(n)} - b_1\|_2 \le \left\| \mathbf{1}_{\{t<U\}} h\left(\frac{t}{U}\right)\left(\left(\frac{I_r^{(n)}}{n}\right)^{\beta} - (UV)^\beta\right)\right\|_2 + C\left\|\frac{(I_1^{(n)})^{\beta-\varepsilon}}{n^\beta}\right\|_2,$$

for some constant $C > 0$. Since h is bounded, the first summand is $O(n^{-1/2})$ just like in (22) above. The second term is trivially bounded by $Cn^{-\varepsilon}$. Overall, we have $\|b_1^{(n)} - b_1\|_2 = O(n^{-\varepsilon})$. Hence, since the coefficients $A_r^{(n)}$ are bounded by one in the operator norm and by distributional properties of $I_1^{(n)}, \dots, I_4^{(n)}$, the first two constraints in Assumption (A1) are satisfied with $R(n) = Cn^{-\varepsilon}$ for a suitable constant $C > 0$, and $\varepsilon > 0$ may still be chosen as small as we want.

Next, we consider L^* in (A1). By dominated convergence we have

$$L^* = 4\mathbf{E}\left[(UV)^{2\beta}(UV)^{-\varepsilon}\right] = \frac{4}{(2\beta - \varepsilon + 1)^2} < 1,$$

for $\varepsilon > 0$ sufficiently small. This completes the verification of (A1).

(A4) PERTURBATION CONDITION. Note that Q_n is piecewise constant: $Q_n(t) = Q_n(s)$ for all s, t if no x-coordinate of the first n points lies between s and t. There are n independent points, the probability that there exists two lying within n^{-3} of each other is at most n^{-1}. So (A4) is satisfied with $r_n = n^{-3}$.

(A5) RATE OF CONVERGENCE. With $r_n = n^{-3}$ and $R_n = Cn^{-\varepsilon}$, we have $R_n = o(\log^{-2} n) = o(\log^{-2}(1/r_n))$ Therefore, the condition on the rate of convergence is satisfied.

4 The limit process

In this section, we prove the existence of a process $Z \in \mathcal{C}[0,1]$, the space of continuous functions from $[0,1]$ into $\mathbb{R}$, that satisfies the distributional fixed point equation (14) and whose mean matches the mean of the rescaled version $Y_n(s)$ of $C_n(s)$. We construct the process Z as the point-wise limit of martingales. We then show that the convergence is actually almost surely uniform, which allows us to conclude that $Z \in \mathcal{C}[0,1]$ with probability one. Figure 2 shows a simulation of the process Z.

We identify the nodes of the infinite quaternary tree with the set of finite words on the alphabet $\{1,2,3,4\}$,

$$\mathcal{T} = \bigcup_{n \ge 0} \{1,2,3,4\}^n.$$

For a node $u \in \mathcal{T}$, we write $|u|$ for its depth, i.e. the distance between u and the root $\varnothing$. The descendants of $u \in \mathcal{T}$ correspond to all the words in $\mathcal{T}$ with prefix u; in particular, the children of u are $u1, \dots, u4$. Let $\{U_v, v \in \mathcal{T}\}$ and $\{V_v, v \in \mathcal{T}\}$ be two independent families of i.i.d. $[0,1]$-uniform random variables.

By $\mathcal{C}_0[0,1]$ we denote the set of continuous functions on the unit interval vanishing at the boundary, i.e. $f(0)=f(1)=0$ for $f\in\mathcal{C}_0[0,1]$. Define the continuous operator $G:(0,1)^2\times\mathcal{C}_0[0,1]^4\to\mathcal{C}_0[0,1]$ by

$$\begin{aligned}G(x,y,f_1,f_2,f_3,f_4)(s)=&\mathbf{1}_{\{s<x\}}\left[(xy)^\beta f_1\left(\frac{s}{x}\right)+(x(1-y))^\beta f_2\left(\frac{s}{x}\right)\right]\\&+\mathbf{1}_{\{s\geq x\}}\left[((1-x)y)^\beta f_3\left(\frac{s-x}{1-x}\right)+((1-x)(1-y))^\beta f_4\left(\frac{s-x}{1-x}\right)\right].\end{aligned}\tag{23}$$

Recall the definition of h in (4). For every node $u\in\mathcal{T}$, let $Z_0^u=h$. Then define recursively

$$Z_{n+1}^u=G(U_u,V_u,Z_n^{u1},Z_n^{u2},Z_n^{u3},Z_n^{u4}).\tag{24}$$

Finally, define $Z_n=Z_n^\varnothing$ to be the value observed at the root of $\mathcal{T}$ when the iteration has been started with h in all the nodes at level n. We will see that for every $s\in[0,1]$, the sequence $(Z_n(s),n\geq 0)$ is a non-negative discrete time martingale; so it converges with probability one to a finite limit.

It will be convenient to have an explicit representation for Z_n. For $s\in[0,1]$, $Z_n(s)$ is the sum of exactly 2^n terms, each one being the contribution of one of the boxes at level n that is cut by the line at s. Let $\{Q_i^n(s),1\leq i\leq 2^n\}$ be the set of rectangles at level n whose first coordinate intersect s. Suppose that the projection of $Q_i^n(s)$ on the first coordinate yields the interval $[\ell_i^n,r_i^n]$. Then

$$Z_n(s)=\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s))^\beta\cdot h\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right),\tag{25}$$

where $\mathrm{Leb}(Q_i^n(s))$ denotes the volume of the rectangle $Q_i^n(s)$. The difference between Z_n and Z_{n+1} only relies in what happens inside the boxes $Q_i^n(s)$: We have

$$Z_{n+1}(s)-Z_n(s)=\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s))^\beta\cdot\left[G(U_i',V_i',h,h,h,h)\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right)-h\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right)\right],\tag{26}$$

where U_i',V_i', $1\leq i\leq 2^n$ are i.i.d. $[0,1]$-uniform random variables. In fact, U_i' and V_i' are some of the variables U_u,V_u for nodes u at level n. Observe that, although the area $\mathrm{Leb}(Q_i^n(s))$ is *not* a product of n independent terms of the form UV because of size-biasing, but U_i',V_i' are in fact *unbiased*, i.e. uniform. Let $\mathscr{F}_n$ denote the σ-algebra generated by $\{U_u,V_u:|u|<n\}$. Then the family $\{U_i',V_i':1\leq i\leq 2^n\}$ is independent of $\mathscr{F}_n$.

So, to prove that $Z_n(s)$ is a martingale, it suffices to prove that, for $1\leq i\leq 2^n$,

$$\mathbf{E}\left[\,G(U_i',V_i',h,h,h,h)\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right)\,\middle|\,\mathscr{F}_n\,\right]=h\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right).$$

Since $U_i',V_i',1\leq i\leq 2^n$ are independent of $\mathscr{F}_n$, this clearly reduces to the following lemma.

Lemma 8. *For the operator G defined in (23) and U,V two independent $[0,1]$-uniform random variables, and any $s\in[0,1]$, we have*

$$\mathbf{E}\left[G(U,V,h,h,h,h)(s)\right]=h(s).$$

Proof. Since V and $1-V$ have the same distribution, we have

$$\mathbf{E}\left[G(U,V,h,h,h,h)(s)\right]=2\mathbf{E}\left[\mathbf{1}_{\{s<U\}}(UV)^\beta h\left(\frac{s}{U}\right)\right]+2\mathbf{E}\left[\mathbf{1}_{\{s\geq U\}}((1-U)V)^\beta h\left(\frac{1-s}{1-U}\right)\right].$$

Similarly, since U and $1-U$ are both uniform, we clearly have

$$\mathbf{E}\left[G(U,V,h,h,h,h)(s)\right]=f(s)+f(1-s),$$

where we wrote $f(s)=2\mathbf{E}\left[\mathbf{1}_{\{s<U\}}(UV)^\beta h(s/U)\right]$. To complete the proof, it suffices to compute $f(s)$. We have

$$\begin{aligned}f(s)=\mathbf{E}\left[\mathbf{1}_{\{s<U\}}(UV)^\beta h\left(\frac{s}{U}\right)\right]&=\frac{2}{\beta+1}\mathbf{E}\left[\mathbf{1}_{\{s<U\}}s^{\beta/2}(U-s)^{\beta/2}\right]\\&=\frac{2}{\beta+1}s^{\beta/2}\int_s^1(x-s)^{\beta/2}dx\\&=\frac{4}{(\beta+1)(\beta+2)}s^{\beta/2}(1-s)^{\beta/2+1}\\&=(1-s)h(s),\end{aligned}$$

where the last line follows since $(\beta+1)(\beta+2)=4$ by definition of β. The result follows readily. $\square$

Our aim is now to prove the following proposition:

Proposition 9. *With probability one Z_n converges uniformly to some continuous limit process Z on $[0,1]$.*

Assume for the moment that there exist constants $a, b \in (0,1)$ and C such that

$$\mathbf{P}\left(\sup_{s\in[0,1]} |Z_{n+1}(s) - Z_n(s)| \geq a^n\right) \leq C \cdot b^n. \tag{27}$$

Then, by the Borel–Cantelli lemma the sequences Z_n is almost surely Cauchy with respect to the supremum norm. Completeness of $(\mathcal{C}[0,1], \|\cdot\|)$ yields the existence of a random process Z with continuous paths such that $Z_n \to Z$ uniformly on $[0,1]$. We now move on to showing that there exist constants a and b such that (27) is satisfied. We start by a bound for a fixed value $s \in [0,1]$. We will then handle the supremum using a sieve of the interval $[0,1]$ by a large enough number of deterministic points.

Lemma 10. *For every $s \in [0,1]$, any $a \in (0,1)$, and any integer n large enough, we have the bound*

$$\mathbf{P}\left(|Z_{n+1}(s) - Z_n(s)| \geq a^n\right) \leq 4(16e\log(1/a))^n.$$

Proof. We use the representation (26). As we have already pointed out earlier (Lemma 8), for every single rectangle $Q_i^n(s)$ at level n, we have

$$\mathbf{E}\left[G(U_i', V_i', h, h, h, h)\left(\frac{s-\ell_i^n}{r_i^n - \ell_i^n}\right) - h\left(\frac{s-\ell_i^n}{r_i^n-\ell_i^n}\right) \,\middle|\, \mathscr{F}_n \right] = 0.$$

Since $h(x) \leq 2^{-\beta}$ for $x \in (0,1)$, conditional on $\mathscr{F}_n$, $Z_{n+1} - Z_n$ is a sum of 2^n *centered*, *bounded* and moreover *independent* terms (but not identically distributed). Moreover, conditional on $\mathscr{F}_n$, the term corresponding to $Q_i^n(s)$ in (26) is bounded by

$$\begin{aligned}\mathrm{Leb}(Q_i^n)^\beta \cdot \|G(U_i', V_i', h, h, h, h) - h\| &\leq \mathrm{Leb}(Q_i^n)^\beta 2\|h\| \\ &= \mathrm{Leb}(Q_i^n)^\beta 2^{1-\beta}.\end{aligned} \tag{28}$$

So when conditioning on $\mathscr{F}_n$, one can bound the variations of $Z_{n+1} - Z_n$ using the Chernoff–Hoeffding inequality [23]. We have

$$\begin{aligned}\mathbf{P}\left(|Z_{n+1}(s) - Z_n(s)| > a^n\right) &= \mathbf{E}\left[\mathbf{P}(|Z_{n+1}(s) - Z_n(s)| > a^n \mid \mathscr{F}_n)\right] \\ &\leq \mathbf{E}\left[2\exp\left(-\frac{a^{2n}}{\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s))^{2\beta}}\right)\right] \\ &\leq 2\exp\left(-a^{-2n}\right) + 2\mathbf{P}\left(\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s))^{2\beta} > a^{4n}\right);\end{aligned} \tag{29}$$

the precise constant in the exponent in the second inequality can be taken to be one since $2/(2^{1-\beta})^2 > 1$.

Now, since $2\beta > 1$ and all the volumes $\mathrm{Leb}(Q_i^n(s))$ are at most one, we have

$$\begin{aligned}\mathbf{P}\left(\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s))^{2\beta} > a^{4n}\right) &\leq \mathbf{P}\left(\sum_{i=1}^{2^n}\mathrm{Leb}(Q_i^n(s)) > a^{4n}\right) \\ &\leq \mathbf{P}\left(W_n > a^{4n}\right),\end{aligned} \tag{30}$$

where W_n denotes the maximum width of any of the 4^n cells at level n. Indeed, the volume occupied by all rectangles $Q_i^n(s)$, $1 \leq i \leq 2^n$ together is at most that of a vertical tube of width W_n. Putting together (29) and (30), it follows that,

$$\begin{aligned}\mathbf{P}\left(|Z_{n+1}(s) - Z_n(s)| \geq a^n\right) &\leq 2\exp(-a^{-2n}) + 2\mathbf{P}\left(W_n > a^{4n}\right) \\ &\leq 2\exp(-a^{-2n}) + 2(16e\log(1/a))^n \\ &\leq 4(16e\log(1/a))^n,\end{aligned}$$

for all n large enough using Lemma 22 from the appendix. $\square$

Now that we have good control on pointwise variations of $Z_{n+1} - Z_n$, we move on to the supremum on $[0,1]$. Consider the set V_n of x-coordinates of the vertical boundaries of all the rectangles at level n. Let $L_n = \inf\{|x-y| : x, y \in V_n\}$. Suppose that $1/\gamma$ is an integer. Then, we have

$$\begin{aligned}
&\sup_{s\in[0,1]} |Z_{n+1}(s) - Z_n(s)| \\
&\le \sup_{1\le i\le \gamma^{-(n+1)}} |Z_{n+1}(i\gamma^{n+1}) - Z_n(i\gamma^{n+1})| + 2 \sup_{m\in\{n,n+1\}} \sup_{|s-t|\le\gamma^{n+1}} |Z_m(s) - Z_m(t)|.
\end{aligned}$$

We first deal with the second term, and suppose that we are on the event that $L_{n+1} \ge (4\gamma)^{n+1}$. Observe that the sieve we used, γ^n, is much finer than the shortest length of a cell at level $n+1$ which is at least L_{n+1}. We use the representation in (25); for $|t-s| \le \gamma^{n+1}$, the two collections $\{Q_i^n(s), 1 \le i \le 2^n\}$ and $\{Q_i^n(t), 1 \le i \le 2^n\}$ differ at most on one cell. We obtain, for any $|s-t| \le \gamma^{n+1}$,

$$\begin{aligned}
|Z_n(s) - Z_n(t)| &\le \sum_{i=1}^{2^n} \mathrm{Leb}(Q_i^n(s))^\beta \cdot \left| h\left(\frac{s-\ell_i^n}{r_i^n - \ell_i^n}\right) - h\left(\frac{t-\ell_i^n}{r_i^n-\ell_i^n}\right)\right| + 2\max_i \mathrm{Leb}(Q_i^n(s))^\beta \\
&\le \sum_{i=1}^{2^n} \mathrm{Leb}(Q_i^n(s))^\beta \cdot 4^{-\beta n} + 2\max_i \mathrm{Leb}(Q_i^n(s))^\beta \\
&\le 3W_n^\beta.
\end{aligned}$$

Here, the second inequality follows from the facts that $|h(t) - h(s)| \le |t-s|^\beta$ for any $s, t \in [0,1]$ and that $L_n \ge (4\gamma)^{n+1}$. The same upper bound is valid for $|Z_{n+1}(s) - Z_{n+1}(t)|$ for $|s-t| \le \gamma^{n+1}$. In particular, it follows by the union bound that, for any $\gamma \in (0,1)$ (with $1/\gamma$ an integer),

$$\begin{aligned}
\mathbf{P}\left(\sup_{s\in[0,1]} |Z_{n+1}(s) - Z_n(s)| \ge 2a^n\right) \le& \gamma^{-n} \sup_{s\in[0,1]} \mathbf{P}\left(|Z_{n+1}(s) - Z_n(s)| \ge a^n\right) \\
&+ \mathbf{P}\left(L_{n+1} < (4\gamma)^{n+1}\right) + \mathbf{P}\left(12W_n^\beta > a^n\right). \qquad (31)
\end{aligned}$$

We are now ready to complete the proof of Proposition 9. From (31) and Lemma 24 from the appendix, we have

$$\mathbf{P}\left(\sup_{s\in[0,1]} |Z_{n+1}(s) - Z_n(s)| \ge 2a^n\right) \le 4(16e\gamma^{-1}\log(1/a))^n + 6\cdot 16^n \gamma^{n/201} + (4e\log(12^{1/n}/a)/\beta)^n,$$

for all $\gamma < \gamma_0/4$ and $n \ge n_0(\gamma, a)$. Now, first choose $a < 1$ sufficiently close to 1 such that we also have $16(e\log(1/a))^{1/202} < 1/4$ and then $\gamma > 0$ such that $1/\gamma$ is an integer and $\gamma^{1/201} \le e\gamma^{-1}\log(1/a)$.

It follows that, for n sufficiently large,

$$\mathbf{P}\left(\sup_{s\in[0,1]} |Z_{n+1}(s) - Z_n(s)| \ge 2a^n\right) \le 11 \cdot 4^{-n}.$$

Increasing $a < 1$ and C ensures that (27) holds with $b = 1/4$ for all $n \ge 1$. The functions $Z_n^1, \dots, Z_n^4$ at the four children of the root are each distributed as Z_{n-1}, and they also converge uniformly to continuous limits denoted $Z^{(1)}, \dots, Z^{(4)}$. The random functions $Z^{(1)}, \dots, Z^{(4)}$ are independent and distributed as Z. Equation (24) and independence imply

$$\begin{aligned}
Z(s) =& \mathbf{1}_{\{s<U\}} \left[(UV)^\beta Z^{(1)}\left(\frac{s}{U}\right) + (U(1-V))^\beta Z^{(2)}\left(\frac{s}{U}\right)\right] \\
&+ \mathbf{1}_{\{s\ge U\}} \left[((1-U)V)^\beta Z^{(3)}\left(\frac{s-U}{1-U}\right) + ((1-U)(1-V))^\beta Z^{(4)}\left(\frac{s-U}{1-U}\right)\right],
\end{aligned}$$

almost surely, considered as random continuous paths. In particular, the distribution of Z solves the distributional fixed-point equation (14).

Finally, we look at the moments of $\|Z_n\| = \sup_{s\in[0,1]} |Z_n(s)|$ and $\|Z\| = \sup_{s\in[0,1]} |Z(s)|$.

Proposition 11. *For every $p \ge 1$, we have $\mathbf{E}[\|Z\|^p] < \infty$ and $\|Z_n - Z\| \to 0$ in L^p.*

Proof. Let $\Delta(x) = \mathbf{P}(\|Z_{n+1} - Z_n\| \geq x)$ and $a < 1, C > 0$ such that (27) is satisfied with $b = 1/4$. Then, by (26) and the upper bound (28), we have

$$\mathbf{E}[\|Z_{n+1} - Z_n\|] = \int_0^\infty \Delta_n(x)dx = \int_0^{a^n} \Delta_n(x)dx + \int_{a^n}^{2^{n+1}} \Delta_n(x)dx. \tag{32}$$

The first summand is at most a^n, the second one at most $C \cdot 2^{-(n-1)}$ by (27). Altogether, there exists $R > 0$ and $0 < q < 1$ with

$$\mathbf{E}[\|Z_{n+1} - Z_n\|] \leq Rq^n$$

for all n. Furthermore, for any $p \in \mathbb{N}$, our proof also provides (27) for a constant $C > 0$ and $b = 4^{-p}$ by increasing the value of a. Therefore, replacing a^n and 2^{n+1} by a^{np} resp. $2^{(n+1)p}$ in (32) shows that the p-th moment of $\|Z_{n+1} - Z_n\|$ is also exponentially small in n for any $p > 1$. Then, since $Z_n = h + \sum_{k=1}^n (Z_k - Z_{k-1})$, using Minkowski's inequality

$$\mathbf{E}[\|Z_n\|^p]^{1/p} \leq \sum_{k=1}^n \mathbf{E}[\|Z_k - Z_{k-1}\|^p]^{1/p} + \|h\|,$$

which is uniformly bounded in n. It follows that $\mathbf{E}[\|Z\|^p] < \infty$ for all $p \geq 1$, and that $\mathbf{E}[\|Z_n - Z\|^p] \to 0$ as $n \to \infty$. □

5 Uniform convergence of the mean

The proof that Assumption (A1) holds for Proposition 7 requires to show convergence of the first moment $n^{-\beta}\mathbf{E}[C_n(s)]$ towards $\mu_1(s) = K_1 h(s)$ uniformly on $[0,1]$. Note that, since $C_n(s)$ is continuous at any fixed $s \in [0,1]$ almost surely, the function $s \to \mathbf{E}[C_n(s)]$ is continuous for any n. Curien and Joseph [6] only show point-wise convergence, and proving uniform convergence requires a good deal of additional arguments. Unfortunately, a good portion of the work consist in tedious tightening of the strategy developed in [6].

Proposition 12. *There exists $\varepsilon > 0$ such that*

$$\sup_{s \in [0,1]} |n^{-\beta}\mathbf{E}[C_n(s)] - \mu_1(s)| = O(n^{-\varepsilon}).$$

In other words, $n^{-\beta}\mathbf{E}[C_n(s)]$ converges uniformly to μ_1 on $[0,1]$ with polynomial rate.

We prove a Poissonized version. Since $C_n(s)$ is increasing in n for every fixed s, the de-Poissonization only relies on routine arguments based on concentration for Poisson random variables and we omit the details. Consider a Poisson point process with unit intensity on $[0,1]^2 \times [0,\infty)$. The first two coordinates represent the location inside the unit square; the third one represents the time of arrival of the point. Let $P_t(s)$ denote the partial match cost for a query at $x = s$ in the quadtree built from the points arrived by time t.

Proposition 13. *There exists $\varepsilon > 0$ such that*

$$\sup_{s \in [0,1]} |t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)| = O(t^{-\varepsilon}).$$

The proof of Proposition 13 relies crucially on two main ingredients: first, a strengthening of the arguments developed by Curien and Joseph [6], and the speed of convergence $\mathbf{E}[C_n(\xi)]$ to $\mathbf{E}[\mu_1(\xi)]$ for a uniform query line ξ, see (2). By symmetry, we write for any $\delta \in (0, 1/2)$

$$\begin{aligned}\sup_{s \in [0,1]} |t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)| &= \sup_{s \in [0,1/2]} \left|t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)\right| \\ &\leq \sup_{s \leq \delta} \left|t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)\right| + \sup_{s \in (\delta,1/2]} \left|t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)\right|. \end{aligned} \tag{33}$$

The two terms in the right hand side above are controlled by the following lemmas.

Lemma 14 (Behavior on the edge)**.** *We have*

$$\sup_{s\le\delta}\left|t^{-\beta}\mathbf{E}[P_t(s)]-\mu_1(s)\right|\le 2^{\beta}\sup_{r\ge t/2} r^{-\beta}\mathbf{E}\left[P_r(\delta)\right]+K_1\delta^{\beta/2}. \tag{34}$$

Lemma 15 (Behavior away from the edge)**.** *There exist constants C_1, C_2, η with $0<\eta<\beta$ and $\gamma\in(0,1)$ such that, for any integer k, and real number $\delta\in(0,1/2)$ we have, for any real number $t>0$,*

$$\sup_{s\in[\delta,1/2]}|t^{-\beta}\mathbf{E}[P_t(s)]-\mu_1(s)|\le C_1\delta^{-1}(1-\gamma)^k+C_2k2^k(\beta-\eta)^{-2k}t^{-\eta}.$$

Before going further, we indicate how these two lemmas imply Proposition 13. By Lemmas 14 and 15, we have for any $\delta\in(0,1/2)$ and natural number $k\ge 0$

$$\sup_{s\in[0,1]}|t^{-\beta}\mathbf{E}[P_t(s)]-\mu_1(s)|\le 3K_1\delta^{\beta/2}+3C_1\delta^{-1}(1-\gamma)^k+5C_2kt^{-\eta}2^k(\beta-\eta)^{-2k}.$$

Choosing $\delta=t^{-\nu}$ and $k=\lfloor\alpha\log t\rfloor$ for $\nu,\alpha>0$ to be determined, we obtain

$$\sup_{s\in[0,1]}|t^{-\beta}\mathbf{E}[P_t(s)]-\mu_1(s)|\le 3K_1t^{-\nu\beta/2}+3C_1t^{\nu}(1-\gamma)^{\alpha\log t-1}+5C_2t^{-\eta}[2/(\beta-\eta)^2]^{\alpha\log t}\alpha\log t.$$

First pick $\alpha>0$ small enough that

$$\alpha\log\left(\frac{2}{(\beta-\eta)^2}\right)<\eta.$$

This α being fixed, choose $\nu>0$ small enough that $\nu+\alpha\log(1-\gamma)<0$. The claim follows.

Since Curien and Joseph [6] prove convergence at any $s\in(0,1)$, it comes as no surprise that the convergence may be strengthened to uniform convergence on compacts of $(0,1)$ by checking carefully the (long) sequence of bounds in [6] (Lemma 15). We provide the details in the appendix for the sake of completeness. The behavior at the edge however (Lemma 14) consists precisely in controlling what happens when the bounds in [6] do not work any longer; this is why we provide here the additional arguments. To deal with the term involving the values of $s\in[0,\delta]$, we relate the value $\mathbf{E}[P_t(s)]$ to $\mathbf{E}[P_t(\delta)]$. The term $\mathbf{E}[P_t(\delta)]$ will then be shown to be small using the pointwise convergence and choosing δ small.

The function $\mu_1(s)=\lim_{t\to\infty}\mathbf{E}[P_t(s)]$ is monotonic for $s\in[0,1/2]$. It seems, at least intuitively, that for any fixed real number $t>0$, $\mathbf{E}[P_t(s)]$ should also be monotonic for $s\in[0,1/2]$, but we were unable to prove it. The following weaker version will be sufficient for our needs.

Proposition 16 (Almost monotonicity)**.** *For any $s<1/2$ and $\varepsilon\in[0,1-2s)$, we have*

$$\mathbf{E}[P_t(s)]\le\mathbf{E}\left[P_{t(1+\varepsilon)}\left(\frac{s+\varepsilon}{1+\varepsilon}\right)\right].$$

The idea underlying Proposition 16 requires to understand what happens to the quadtree upon considering a larger point set. For a finite point set $\mathcal{P}\subset[a,b]\times[0,1]\times[0,\infty)$, we let $V(\mathcal{P})$ and $H(\mathcal{P})$ denote, respectively, the set of vertical and horizontal line segments of the quadtree built from $\mathcal{P}$.

Lemma 17. *Let $\mathcal{P}=\{p_1,\dots,p_n\}$ be a set of points with $p_i=(x_i,y_i,t_i)\in[a_2,a_3]\times[0,1]\times[0,\infty)$ ordered by their t coordinate, i.e. $t_i\le t_{i+1}$. Additionally we assume $\mathcal{P}$ to be in general position, meaning that all x-coordinates are pairwise different and the same holds true for the y and t coordinates. Furthermore let $\mathcal{Q}=\{p'_1,\dots,p'_m\}\subseteq[a_1,a_2]\times[0,1]\times[0,\infty)$ with $p'_i=(x'_i,y'_i,t'_i)$ again ordered according to their third coordinate such that $\mathcal{P}\cup\mathcal{Q}\subseteq[a_1,a_3]\times[0,1]\times[0,\infty)$ is again in general position. Then we have*

$$H(\mathcal{P}\cup\mathcal{Q})\supset H(\mathcal{P})\qquad\text{and}\qquad V(\mathcal{P}\cup\mathcal{Q})\subset V(\mathcal{P}).$$

Proof. We assume for a contradiction that the assertion is wrong and focus on the case that $H(\mathcal{P})\not\subset H(\mathcal{P}\cup\mathcal{Q})$; the other case is handled analogously. Let i_1 be the index of the "first" point in $\mathcal{P}$ such that the horizontal line of p_{i_1} is shorter (at least on the right or left side of the point) in the quadtree built from $\mathcal{P}\cup\mathcal{Q}$ than it is in the one built from $\mathcal{P}$. Here, first refers to the time coordinate t. Now, by construction there must be an index i_2 such that the vertical line of p_{i_2} blocks the horizontal line of p_{i_1} in $\mathcal{P}\cup\mathcal{Q}$ but not in $\mathcal{P}$. We again choose i_2 such that t_{i_2} is minimal with this property; by construction $t_{i_2}<t_{i_1}$. Repeating the argument gives the existence of an index i_3 and a point p_{i_3} whose horizontal line blocks the vertical line of p_{i_2} in $\mathcal{P}$ but not in $\mathcal{P}\cup\mathcal{Q}$ with $t_{i_3}<t_{i_2}$. This obviously contradicts the choice of i_1. ☐

Proof of Proposition 16. Consider the unit square $[0,1]^2$ and the extended box $[-\varepsilon, 1] \times [0,1]$, and a single Poisson point process on $[-\varepsilon, 1] \times [0,1] \times [0,t]$ with unit intensity. Write $P_t^\varepsilon(s)$ for the number of (horizontal) lines intersecting $\{x = s\}$ in the quadtree formed by the all the points. Similarly, let $P_t(s) = P_t^0(s)$ be the corresponding quantity when the quadtree is formed using only the points falling inside $[0,1]^2$. Then, for this coupling, we have by Lemma 17,

$$P_t(s) \le P_t^\varepsilon(s) \overset{d}{=} P_{t(1+\varepsilon)}\left(\frac{s+\varepsilon}{1+\varepsilon}\right).$$

Taking expectations completes the proof. □

Proof of Lemma 14. We use Proposition 16 to relate $\mathbf{E}[P_t(s)]$ to $\mathbf{E}[P_{t'}(\delta)]$ for some t'. Choosing $\varepsilon = (\delta - s)/(1-\delta)$ yields $t' = t(1-s)/(1-\delta) \le t(1-\delta)^{-1}$. Thus, for any $\delta \in (0, 1/2)$ and $t > 0$ we have

$$\begin{aligned}\sup_{s\le\delta} \left|t^{-\beta}\mathbf{E}[P_t(s)] - \mu_1(s)\right| &\le \sup_{s\le\delta} t^{-\beta}\mathbf{E}[P_t(s)] + \mu_1(\delta) \\ &\le \sup_{s\le\delta} t^{-\beta}\mathbf{E}[P_{t'}(\delta)] + \mu_1(\delta) \\ &\le t^{-\beta}\mathbf{E}[P_{t/(1-\delta)}(\delta)] + \mu_1(\delta) \\ &\le (1-\delta)^{-\beta} \sup_{r\ge t/2} r^{-\beta}\mathbf{E}[P_r(\delta)] + \mu_1(\delta).\end{aligned}$$

This completes the proof since $\delta \le \frac{1}{2}$ and $\mu_1(s) \le K_1\delta^{\beta/2}$. □

6 Moments and supremum: Proofs of Theorems 4, 5 and Corollary 6

Our main result implies the convergence of the second moment of the discrete towards that of the limit process. This section is devoted to identifying this limit, in particular it provides an explicit expression for the limit variance.

We first focus on the moments. The definition of the process $Z(s)$ implies that the second moment $\mu_2(s) = \mathbf{E}[Z(s)^2]$ satisfies an integral equation. We have

$$\begin{aligned}\mu_2(s) = \mathbf{E}[Z(s)^2] =& 2\mathbf{E}\left[Y^{2\beta}\right]\left\{\int_s^1 x^{2\beta}\cdot\mu_2\left(\frac{s}{x}\right)dx + \int_0^s (1-x)^{2\beta}\cdot\mu_2\left(\frac{1-s}{1-x}\right)dx\right\} \\ &+ 2\mathbf{E}\left[[Y(1-Y)]^\beta\right]\cdot\left\{\int_s^1 x^{2\beta}h\left(\frac{s}{x}\right)^2 dx + \int_0^s (1-x)^{2\beta}h\left(\frac{1-s}{1-x}\right)^2 dx\right\}.\end{aligned}$$

It now follows that μ_2 satisfies the following integral equation

$$\mu_2(s) = \frac{2}{2\beta+1}\left\{\int_s^1 x^{2\beta}\mu_2\left(\frac{s}{x}\right)dx + \int_0^s (1-x)^{2\beta}\mu_2\left(\frac{1-s}{1-x}\right)dx\right\} + 2B(\beta+1,\beta+1)\cdot\frac{h^2(s)}{\beta+1}.$$

One easily verifies that the function f given by $f(s) = c_2h^2(s)$ solves the above equation when c_2 is given by

$$c_2 = 2\mathrm{B}(\beta+1,\beta+1)\frac{2\beta+1}{3(1-\beta)}. \tag{35}$$

In order to show that $\mu_2 = c_2h(s)^2$, it now suffices to prove that the integral equation satisfied by μ_2 admits a unique solution in a suitable function space. To this aim, we show that the map K defined below is a contraction for the supremum norm:

$$\begin{aligned}Kf(s) =& \frac{2}{2\beta+1}\left\{\int_s^1 x^{2\beta}f\left(\frac{s}{x}\right)dx + \int_0^s (1-x)^{2\beta}f\left(\frac{1-s}{1-x}\right)dx\right\} \\ &+ 2B(\beta+1,\beta+1)\frac{h(s)^2}{\beta+1}.\end{aligned} \tag{36}$$

For any two functions f and g, measurable and bounded on $[0,1]$, we have

$$\begin{aligned}
&\|Kf - Kg\|_\infty \\
&\quad = \frac{2}{2\beta+1} \sup_{s\in[0,1]} \left| \int_s^1 x^{2\beta} \left(f\left(\frac{s}{x}\right) - g\left(\frac{s}{x}\right)\right) dx + \int_0^s (1-x)^{2\beta} \left(f\left(\frac{1-s}{1-x}\right) - g\left(\frac{1-s}{1-x}\right)\right) dx \right| \\
&\quad \leq \frac{2}{2\beta+1} \left(\sup_{s\in[0,1]} \left\{ \int_s^1 x^{2\beta} dx \right\} + \sup_{s\in[0,1]} \left\{ \int_0^s (1-x)^{2\beta} dx \right\} \right) \|f-g\|_\infty \\
&\quad = \frac{4}{(2\beta+1)^2} \|f-g\|_\infty .
\end{aligned}$$

Since $2\beta + 1 > 2$, the operator K is a contraction on the set of measurable and bounded functions on $[0,1]$ equipped with the supremum norm. Banach fixed point theorem then ensures that the fixed point is unique, which shows that indeed $\mathbf{E}[Z(s)^2] = c_2 h^2(s)$. Then, $K_2 = c_2 - 1$ and one obtains easily the expression for $\mathbf{Var}(Z(\xi))$ in (7) by integration.

Analogously one shows that the m-th moment of $Z(s)$ is of the form $c_m h(s)^m$ where c_m solves (11). The Lipschitz constant of the corresponding operator in (36) is $4/(\beta m + 1)^2$, hence again smaller than one. This immediately implies that $(c_m)_{m\geq 1}$ are the moments of $Z(s)/h(s)$, independently of s.

Furthermore, there is only one distribution with these moments. We let Ψ denote the corresponding random variable. To prove this, we show that there exists a constant $A_1 > 0$ such that

$$c_m \leq A_1^m m^m, \qquad m \geq 1, \tag{37}$$

which completes the proof of the proposition by the Carleman condition [see, e.g., 14, p. 228]

Suppose that (37) is satisfied for all $m < m_0$. By Stirling's formula, there exists a constant A_2 such that for all $m \geq 1$ and $1 \leq \ell < m$

$$\binom{m}{\ell} \mathrm{B}(\beta\ell + 1, \beta(m-\ell)+1) \leq \frac{A_2}{m} \left(\frac{\ell^\ell (m-\ell)^{m-\ell}}{m^m} \right)^{\beta - 1}.$$

Next, the prefactor in (11) is of order $1/m$, hence bounded by A_3/m for some $A_3 > 0$ and all $m > 1$. Using this, the induction hypothesis and $x^x(1-x)^{1-x} \leq 1$ for all $x \in [0,1]$ it follows that

$$\begin{aligned}
c_{m_0} &\leq \frac{A_2 A_3}{m_0^2} \sum_{\ell=1}^{m_0-1} \left(\ell^\ell (m_0-\ell)^{m_0-\ell} \right)^{\beta-1} m_0^{m_0(1-\beta)} c_\ell c_{m_0-\ell} \\
&\leq \frac{A_1^{m_0} A_2 A_3}{m_0^2} \sum_{\ell=1}^{m_0-1} m_0^{\beta m_0} m_0^{m_0(1-\beta)} \\
&\leq A_1^{m_0} m_0^{m_0},
\end{aligned}$$

if m_0 is chosen large enough. Finally, it is easy to see that any solution of (10) with unit mean and finite second moment has finite moments of all orders. Thus, its moments also satisfy (11) and it must coincide with Ψ in distribution.

We now consider the supremum $S_n = \sup_{s\in(0,1)} C_n(s)$. The uniform convergence of $n^{-\beta} C_n$ directly implies, as $n \to \infty$,

$$\bar{S}_n := \frac{S_n}{K_1 n^\beta} \to S$$

in distribution with $S = \sup_{t\in[0,1]} Z(t)$ where Z is the process constructed in Section 4. The results obtained so far yield that, stochastically

$$S \leq \left((UV)^\beta S^{(1)} + (U(1-V))^\beta S^{(2)} \right) \vee \left(((1-U)V)^\beta S^{(3)} + ((1-U)(1-V))^\beta S^{(4)} \right), \tag{38}$$

where $S^{(1)}, \ldots, S^{(4)}$ are independent copies of S, also independent of (U,V) which are themselves independent and uniform on $[0,1]$. To complete the proof of Theorem 4, it remains to prove that, for all m, $\mathbf{E}[S^m] < \infty$ and that $\mathbf{E}[\bar{S}_n^m] \to \mathbf{E}[S^m]$, as $n \to \infty$. Theorem 12 and Corollary 21 in [32] provide uniform integrability of $\bar{S}_n^2$. It follows that $\bar{S}_n$ is bounded in L^2 and hence also in L^1. For higher moments, we proceed by induction. Let B_1 be such that $\mathbf{E}[\bar{S}_n^m] \le B_1$ for all $m < m_0$ and $n \ge 1$ with $m_0 \ge 2$. Furthermore, choose B_2 such that $\mathbf{E}[\bar{S}_n^{m_0}] \le B_2$ for all $n < n_0$. Then, the recurrence for $C_n(t)$ yields

$$\begin{aligned}\mathbf{E}\left[\bar{S}_{n_0}^{m_0}\right] \le & \mathbf{E}\left[\left(\left(\frac{I_1^{(n)}}{n}\right)^{\beta}\bar{S}^{(1)}_{I_1^{(n)}} + \left(\frac{I_2^{(n)}}{n}\right)^{\beta}\bar{S}^{(1)}_{I_2^{(n)}}\right)^{m_0}\right] \\ & + \mathbf{E}\left[\left(\left(\frac{I_3^{(n)}}{n}\right)^{\beta}\bar{S}^{(3)}_{I_3^{(n)}} + \left(\frac{I_4^{(n)}}{n}\right)^{\beta}\bar{S}^{(4)}_{I_4^{(n)}}\right)^{m_0}\right] \\ \le & 4^{m_0}B_1^2 + 4B_2\mathbf{E}\left[\left(\frac{I_1^{(n)}}{n}\right)^{\beta m_0}\right].\end{aligned}$$

Note that, as $n \to \infty$, we have $\mathbf{E}[(I_1^{(n)}/n)^{\beta m_0}] \to \mathbf{E}\left[(UV)^{\beta m_0}\right] = (\beta m_0 + 1)^{-2}$, thus choosing n_0 and B_2 appropriately we have $\mathbf{E}[\bar{S}_{n_0}^{m_0}] \le B_2$ since $m_0 \ge 2$. This shows that $\bar{S}_n$ is bounded in L^{m_0}, and the assertion follows.

7 Partial match queries in random 2-d trees

7.1 2-d trees: constructions and recursions

The random 2-d tree was introduced by Bentley [1] and is used to store two-dimensional data just as the two-dimensional quadtree. It is also called two-dimensional binary search tree since it is binary and mimics the construction rule of binary search tree for two-dimensional data. Our aim in this section is to introduce 2-d trees, and extend to 2-d trees the results for partial match queries in quadtrees we obtained in the previous sections. All the results can be transferred (convergence as a process, convergence of all moments at one or multiple points, convergence of the supremum in distribution and for all moments); we will mainly state the forms of the theorems for 2-d trees, and focus on the points that deserve some verifications.

CONSTRUCTION OF 2-D TREES. The data are partitioned recursively, as in quadtrees, but the splits are only binary; since the data is two-dimensional, one alternates between vertical and horizontal splits, depending on the parity of the level in the tree. More precisely, consider a point sequence $p_1, p_2, \dots, p_n \in [0,1]^2$. As we build the tree, regions are associated to each node. Initially, the root is associated with the entire square $[0,1]^2$. The first item p_1 is stored at the root, and splits *vertically* the unit square in two rectangles, which are associated with the two children of the root. More generally, when i points have already been inserted, the tree has i internal nodes, and $i+1$ (lower level) regions associated to the external nodes and forming a partition of the square $[0,1]^2$. When point p_{i+1} is stored in the node, say u, corresponding to the region it fall in, divides the region in two sub-rectangles that are associated to the two children of u, which become external nodes; that last partition step depends on the parity of the depth of u in the tree: if it is odd we partition horizontally, if it is even we partition vertically. See Figure 3. (Of course, one could start at the root with a horizontal split.)

PARTIAL MATCH QUERIES. From now on, we assume that data consists of a set of independent random points, uniformly distributed on the unit square. Unlike in the case of quadtrees, the *direction* of a partial match query line with respect to the direction of the root does matter. Let $C_n^{=}(t)$ and $C_n^{\perp}(t)$ denote the number of nodes visited by a partial match for a query at position $t \in [0,1]$ when the directions of the split at the root and the query are parallel and perpendicular, respectively. Subsequently, we will analyze both quantities synchronously as far as possible. We will always consider directions with respect to the query line, and although some of the expressions (for the sizes of the regions for instance) will be symmetric, we keep them distinct for the sake of clarity. (We also assume without loss of generality that the query line is always vertical, and that the direction of the cut at the root may change.)

As in a quadtree, a node is visited by a partial match query if and only if it is inserted in a subregion that intersects the query line. Unfortunately, these nodes are not easily identifiable *after* the insertion of n points; the value of the quantity $C_n^{=}(s)$ is obtained by adding twice the number of lines intersecting the query line at s and the number of boxes that are intersected by the query line and will have their next split perpendicular to the query line (that is, the depth of the corresponding external nodes in the tree have odd parity).

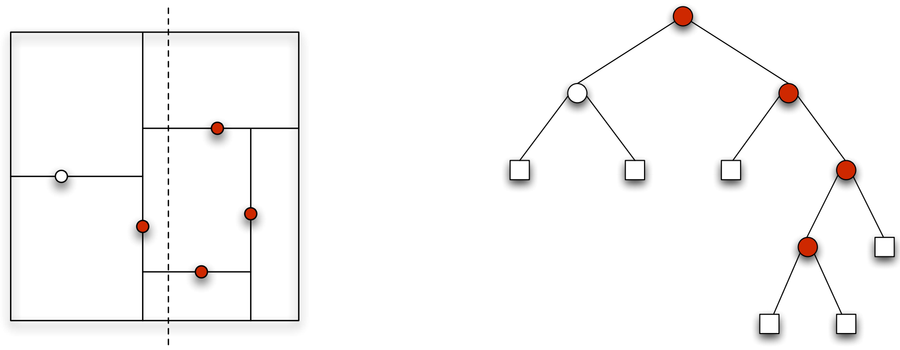

Figure 3: An example of a 2-d tree is shown: on the left, the partition of $[0,1]^2$ induced by the points; on the right, the corresponding binary tree. In red, the nodes visited when performing the partial match query at materialized by the dashed vertical line.

RECURSIVE DECOMPOSITIONS. Let (U, V) be the first point which partitions the unit square. By construction, since the directions of the partitioning lines alternate, both processes $C_n^{=}(t)$ and $C_n^{\perp}(t)$ are coupled: when the query line is perpendicular to the split direction, the recursive search occur in both child sub-regions whose sizes we denote by N_n and S_n, and we have

$$C_n^{\perp}(s) \stackrel{d}{=} 1 + C_{N_n}^{(=,1)}(s) + C_{S_n}^{(=,2)}(s); \tag{39}$$

when the query line and the first split at the root are parallel, only one of the sub-regions (of sizes L_n and R_n) is recursively visited and we have

$$C_n^{\perp}(s) \stackrel{d}{=} 1 + \mathbf{1}_{\{s<U\}} C_{L_n}^{(=,1)}\left(\frac{s}{U}\right) + \mathbf{1}_{\{s\geq U\}} C_{R_n}^{(=,2)}\left(\frac{s-U}{1-U}\right). \tag{40}$$

Here $(C_n^{(=,1)})_{n\geq 0}, (C_n^{(=,2)})_{n\geq 0}$ are independent copies of $(C_n^{=})_{n\geq 0}$, independent of (N_n, S_n) in (39) and $(C_n^{(\perp,1)})_{n\geq 0}, (C^{(\perp,2)})_{n\geq 0}$ are independent copies of $(C_n^{\perp})_{n\geq 0}$, independent of (L_n, R_n) in (40). Moreover, here and in the following distributional recurrences and fixed-point equations involving a parameter $s \in [0,1]$ are to be understood on the level of càdlàg or continuous functions unless stated otherwise.

As in the case of partial match in random quadtrees, the expected value at a random uniform query line ξ, independent of the tree is of order n^β for the same constant β defined in (1), and we have

$$\mathbf{E}[C_n^{=}(\xi)] \sim \kappa_{=} n^\beta, \qquad \mathbf{E}[C_n^{\perp}(\xi)] \sim \kappa_{\perp} n^\beta,$$

for some constants $\kappa_{=} > 0, \kappa_{\perp} > 0$. This was first proved by Flajolet and Puech [17]. A more detailed analysis by Chern and Hwang [5] shows that

$$\mathbf{E}[C_n^{=}(\xi)] = \kappa_{=} n^\beta - 2 + O(n^{\beta-1}), \qquad \kappa_{=} = \frac{13(3-5\beta)}{4} \cdot \frac{\Gamma(2\beta+2))}{\Gamma(\beta+1)^3}, \tag{41}$$

$$\mathbf{E}[C_n^{\perp}(\xi)] = \kappa_{\perp} n^\beta - 3 + O(n^{\beta-1}), \qquad \kappa_{\perp} = \frac{13(2\beta-1)}{2} \cdot \frac{\Gamma(2\beta+2))}{\Gamma(\beta+1)^3}. \tag{42}$$

Observe that $\kappa_{=} = \frac{1}{2} 13(3-5\beta)\kappa$ and $\kappa_{\perp} = 13(2\beta-1)\kappa$, where κ is the leading constant for $\mathbf{E}[C_n(\xi)]$ in the case of quadtrees defined in (1).

HOMOGENEOUS RECURSIVE RELATIONS AND LIMIT BEHAVIOUR. For our purposes, and although it yields more complex expressions, it is more convenient to expand the recursion one more level to obtain recursive relations that only involve quantities of the same type, only $(C_n^{=})_{n\geq 0}$ or only $(C_n^{\perp})_{n\geq 0}$: each one of the first two sub-region at the root is eventually split, and this gives rise to a partition into four regions at level two of the tree. Let (U_ℓ, V_ℓ) and (U_r, V_r) be respectively the first points on each side (left and right) of the first cut, when it is parallel to the query line. Let also (U_u, V_u) and (U_d, V_d) be the first points on each side of the cut (up and down) when it is perpendicular to the query line. Note that U, V_ℓ, V_r are independent and uniform on $[0,1]$, and so are V, U_u and U_d.

Let $I_{=,1}^{(n)}, \ldots, I_{=,4}^{(n)}$, and $I_{\perp,1}^{(n)}, \ldots, I_{\perp,4}^{(n)}$ denote the number of data points falling in these regions when the root and the query line are parallel and perpendicular respectively. The distributions of $I_{=,1}^{(n)}, \ldots, I_{=,4}^{(n)}$ on the one hand, and $I_{\perp,1}^{(n)}, \ldots, I_{\perp,4}^{(n)}$ on the other hand are slightly more involved than in the case of quadtrees. One has e.g. given the values of U, V_ℓ, V_r it holds

$$I_{=,1}^{(n)} \stackrel{d}{=} \mathrm{Bin}((\mathrm{Bin}(n-1;U)-1)_+, V_\ell)$$

and given V, U_d, U_u

$$I_{\perp,1}^{(n)} \stackrel{d}{=} \mathrm{Bin}((\mathrm{Bin}(n-1;V)-1)_+, U_d)$$

where the inner and outer binomials are independent. Analogous expressions hold true for the remaining quantities.

Substituting (39) and (40) into each other gives

$$\begin{aligned} C_n^{=}(s) \stackrel{d}{=} 1 + \mathbf{1}_{\{s<U\}} &\left[\mathbf{1}_{\{L_n>0\}} + C^{(=,1)}_{I_{=,1}^{(n)}}\left(\frac{s}{U}\right) + C^{(=,2)}_{I_{=,2}^{(n)}}\left(\frac{s}{U}\right)\right] \\ + \mathbf{1}_{\{s\geq U\}} &\left[\mathbf{1}_{\{R_n>0\}} + C^{(=,3)}_{I_{=,3}^{(n)}}\left(\frac{s-U}{1-U}\right) + C^{(=,4)}_{I_{=,4}^{(n)}}\left(\frac{s-U}{1-U}\right)\right] \end{aligned} \tag{43}$$

and

$$\begin{aligned} C_n^{\perp}(s) \stackrel{d}{=} 1 + \mathbf{1}_{\{S_n>0\}} + \mathbf{1}_{\{N_n>0\}} &+ \mathbf{1}_{\{s<U_d\}} C^{(\perp,1)}_{I_{\perp,1}^{(n)}}\left(\frac{s}{U_d}\right) + \mathbf{1}_{\{s<U_u\}} C^{(\perp,2)}_{I_{\perp,2}^{(n)}}\left(\frac{s}{U_u}\right) \\ &+ \mathbf{1}_{\{s\geq U_d\}} C^{(\perp,3)}_{I_{\perp,3}^{(n)}}\left(\frac{s-U_d}{1-U_d}\right) + \mathbf{1}_{\{s\geq U_u\}} C^{(\perp,4)}_{I_{\perp,4}^{(n)}}\left(\frac{s-U_u}{1-U_u}\right) \end{aligned} \tag{44}$$

where $(C_n^{(=,i)})_{n\geq 0}$, $i = 1, \ldots, 4$, are independent copies of $(C_n^{=})_{n\geq 0}$, which are also independent of the family $(U, I_{=,1}^{(n)}, I_{=,2}^{(n)}, I_{=,3}^{(n)}, I_{=,4}^{(n)})$ in (43), and $(C_n^{(\perp,i)})_{n\geq 0}$, $i = 1, \ldots, 4$, are independent copies of $(C_n^{\perp})_{n\geq 0}$, which are also independent of $(U_d, U_u, I_{\perp,1}^{(n)}, I_{\perp,2}^{(n)}, I_{\perp,3}^{(n)}, I_{\perp,4}^{(n)})$ in (44). Asymptotically, any limit $Z^{=}(s)$ of $n^{-\beta} C_n^{=}(s)$ should satisfy the following fixed-point equation

$$\begin{aligned} Z^{=}(s) \stackrel{d}{=} &\mathbf{1}_{\{s<U\}} \left[(UV_\ell)^\beta Z^{(=,1)}\left(\frac{s}{U}\right) + (U(1-V_\ell))^\beta Z^{(=,2)}\left(\frac{s}{U}\right)\right] \\ &+ \mathbf{1}_{\{s\geq U\}} \left[((1-U)V_r)^\beta Z^{(=,3)}\left(\frac{s-U}{1-U}\right) + ((1-U)(1-V_r))^\beta Z^{(=,4)}\left(\frac{s-U}{1-U}\right)\right], \end{aligned} \tag{45}$$

where $Z^{(=,i)}$, $i = 1, \ldots, 4$, are independent copies of $Z^{=}$, independent of (U, V_ℓ, V_r). Likewise any limit of $n^{-\beta} C_n^{\perp}(s)$ should satisfy

$$\begin{aligned} Z^{\perp}(s) \stackrel{d}{=} &\mathbf{1}_{\{s<U_d\}} (U_d V)^\beta Z^{(\perp,1)}\left(\frac{s}{U_d}\right) + \mathbf{1}_{\{s<U_u\}} (U_u(1-V))^\beta Z^{(\perp,2)}\left(\frac{s}{U_u}\right) \\ &+ \mathbf{1}_{\{s\geq U_d\}} ((1-U_d)V)^\beta Z^{(\perp,3)}\left(\frac{s-U_d}{1-U_d}\right) \\ &+ \mathbf{1}_{\{s\geq U_u\}} ((1-U_u)(1-V))^\beta Z^{(\perp,4)}\left(\frac{s-U_u}{1-U_u}\right), \end{aligned} \tag{46}$$

where $Z^{(\perp,i)}$, $i = 1, \ldots, 4$, are independent copies of $Z^{\perp}$, independent of (U_d, U_u, V). Moreover, according to (39) and (40), we expect a connection between these two limits. This will be stated in the first result of the next section and always allows us to focus on $C_n^{=}(s)$ first. Result for $C_n^{\perp}$ can then be deduced easily afterwards.

7.2 About the conditions to use the contraction argument

EXISTENCE OF CONTINUOUS LIMIT PROCESSES. As in the case of quadtrees, one of the first steps consists in showing the existence of the limit processes $Z^{\perp}$ and $Z^{=}$.

Proposition 18. *There exist two random continuous processes $Z^=, Z^\perp$ with $\mathbf{E}[Z^=(s)] = \mathbf{E}[Z^H(s)] = h(s)$, finite absolute moments of all orders such that $Z^=$ satisfies (45) and $Z^\perp$ satisfies (46). The laws of $Z^=$ and $Z^\perp$ are both unique under these constraints. Additionally,*

- $$\frac{2}{\beta+1} Z^\perp(s) \stackrel{d}{=} V^\beta Z^{(=,1)}(s) + (1-V)^\beta Z^{(=,2)}(s) \tag{47}$$

  and
  $$\frac{\beta+1}{2} Z^=(s) \stackrel{d}{=} \mathbf{1}_{\{s<U\}} U^\beta Z^{(\perp,1)}\left(\frac{s}{U}\right) + \mathbf{1}_{\{s\ge U\}}(1-U)^\beta Z^{(\perp,2)}\left(\frac{s-U}{1-U}\right).$$

- *for every fixed $s \in [0,1]$, $Z^=(s)$ is distributed like $Z(s)$ where Z is the process constructed in Section 4. In particular,* $\mathbf{Var}\,(Z^=(s))$ *is given in* (8) *and* $\mathbf{Var}\,\left(Z^\perp(s)\right) = K_2^\perp h^2(s)$*, where*

  $$K_2^\perp = \left(\frac{2c_2}{2\beta+1}\left(\frac{\beta+1}{2}\right)^2 + 2\mathrm{B}\left(\beta+1,\beta+1\right)\left(\frac{\beta+1}{2}\right)^2 - 1\right), \tag{48}$$

  and c_2 is defined in (35).

- *if ξ is uniform on $[0,1]$ and independent of $Z^=, Z^\perp$, then* $\mathbf{Var}\,(Z^=(\xi)) = \mathbf{Var}\,(Z(\xi))$ *and*

  $$\begin{aligned}\mathbf{Var}\,\left(Z^\perp(\xi)\right) = K_3^\perp = &\left(\frac{2c_2}{2\beta+1} + 2\mathrm{B}\left(\beta+1,\beta+1\right)\right)\left(\frac{\beta+1}{2}\right)^2 \mathrm{B}(\beta+1,\beta+1) \\ &- \left(\mathrm{B}\left(\frac{\beta}{2}+1,\frac{\beta}{2}+1\right)\right)^2.\end{aligned} \tag{49}$$

Proof. The fixed-point equation (45) is very similar to that in (14), and we use the approach that has proved fruitful in Section 4. More precisely, the construction of $Z(s)$ slightly modified to $Z^=(s)$. Define the operator $G^= : [0,1]^3 \times \mathcal{C}[0,1]^4 \to \mathcal{C}[0,1]$ by

$$\begin{aligned}G^=(x,y,z,f_1,f_2,f_3,f_4)(s) =& \mathbf{1}_{\{s<x\}}\left[(xy)^\beta f_1\left(\frac{s}{x}\right) + (x(1-y))^\beta f_2\left(\frac{s}{x}\right)\right] \\ &+ \mathbf{1}_{\{s\ge x\}}\left[((1-x)z)^\beta f_3\left(\frac{s-x}{1-x}\right) + ((1-x)(1-z))^\beta f_4\left(\frac{s-x}{1-x}\right)\right].\end{aligned}$$

Then let (as in Section 4)

$$Z_{n+1}^{=,u} = G^=(U_u, V_u, W_u, Z_n^{=,u1}, Z_n^{=,u2}, Z_n^{=,u3}, Z_n^{=,u4}), \qquad Z_0^{=,u} = h(s),$$

for all $u \in \mathcal{T}$, where $\{U_v, v \in \mathcal{T}\}, \{V_v, v \in \mathcal{T}\}$ and $\{W_v, v \in \mathcal{T}\}$ are three independent families of i.i.d. $[0,1]$-uniform random variables. Lemma 10 remains true for $Z_n^= := Z_n^{=,\emptyset}$ since $W_n^=$ equals W_n in distribution where W_n appears in (30). Since also $L_n^=$ and L_n (appearing in Lemma 24) coincide in distribution, (27) holds true for $Z_n^=$ and therefore Proposition 9 remains valid. The existence of all moments of $\sup_{s\in[0,1]} Z^=(s)$ follows in the same way. Finally, note that $Z_n^=(s)$ is distributed as $Z_n(s)$ for all fixed n, s, hence the one-dimensional distributions of $Z^=$ and Z coincide. It is now easy to see that $Z^\perp$ defined by (47) solves (46). The uniqueness of $Z^=(s)$ (resp. $Z^=(s)$) follows by contraction with respect to the ζ_2 metric, compare Lemma 18 in [32]. Finally, the variance of $Z^\perp(s)$ can be computed as in Section 6 but it is much easier to use (47), we omit the calculations. □

UNIFORM CONVERGENCE OF THE MEAN. Comparing construction and recurrence for partial match queries in 2-d trees and quadtrees it seems very likely that this quantities are not only of the same asymptotic order in the case of a uniform query but also closely related for fixed $s \in [0,1]$ and $n \in \mathbb{N}$. This can be formalized by the following lemma:

Lemma 19. *For any $s \in [0,1]$ and $n \in \mathbb{N}$ we have*

$$\frac{1}{5}\mathbf{E}[C_n(s)] \le \mathbf{E}[C_n^=(s)] \le 2\mathbf{E}[C_n(s)].$$

Proof. We prove both bounds by induction on n using the recursive decompositions (13), (43). Both inequalities are obviously true for $n = 0, 1$. Assume that the assertions were true for all $m \le n-1$ and $s \in [0,1]$. We start with the upper bound which is easier. By (43), we have

$$\mathbf{E}\left[C_n^{=}(s)\right] \le 2 + \mathbf{E}\left[\mathbf{1}_{\{s<U\}}\left[C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right) + C^{(=,2)}_{I^{(n)}_{=,2}}\left(\frac{s}{U}\right)\right]\right] + \mathbf{E}\left[\mathbf{1}_{\{s\ge U\}}\left[C^{(=,3)}_{I^{(n)}_{=,3}}\left(\frac{s-U}{1-U}\right) + C^{(=,4)}_{I^{(n)}_{=,4}}\left(\frac{s-U}{1-U}\right)\right]\right].$$

Hence, it suffices to show that

$$\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right] \le 2\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{1}}\left(\frac{s}{U}\right)\right].$$

This can be done in two steps. First, by conditioning on $I^{(n)}_{=,1}$ and U, using the induction hypothesis, we have

$$\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right] \le 2\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right].$$

Finally, conditioning on U, $I^{(n)}_{=,1}$ is stochastically smaller than $I^{(n)}_1$ which gives

$$\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right] \le 2\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{1}}\left(\frac{s}{U}\right)\right].$$

by monotonicity of $n \to \mathbf{E}\left[C_n(s)\right]$. For the lower bound, note that

$$\mathbf{E}\left[C_n^{=}(s)\right] \ge 1 + \mathbf{E}\left[\mathbf{1}_{\{s<U\}}\left[C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right) + C^{(=,2)}_{I^{(n)}_{=,2}}\left(\frac{s}{U}\right)\right]\right] + \mathbf{E}\left[\mathbf{1}_{\{s\ge U\}}\left[C^{(=,3)}_{I^{(n)}_{=,3}}\left(\frac{s-U}{1-U}\right) + C^{(=,4)}_{I^{(n)}_{=,4}}\left(\frac{s-U}{1-U}\right)\right]\right].$$

Therefore, it is enough to prove

$$\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right] \ge \frac{1}{5}\left(\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{1}}\left(\frac{s}{U}\right)\right] - 1\right).$$

This can be done as for the upper bound. First, by the induction hypothesis, we have

$$\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(=,1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right] \ge \frac{1}{5}\mathbf{E}\left[\mathbf{1}_{\{s<U\}}C^{(1)}_{I^{(n)}_{=,1}}\left(\frac{s}{U}\right)\right].$$

The result follows as for the upper bound by the fact that $I^{(n)}_{=,1}$ is stochastically larger than $(I^{(n)}_1 - 1)^+$ and $C^{(1)}_{(I^{(n)}_1-1)^+} \ge C^{(1)}_{I^{(n)}_1} - 1$. ◻

Recalling (41) and (42), it is natural to introduce the constants

$$K_1^{=} = \frac{\kappa_{=}}{\mathrm{B}\left(\frac{\beta}{2}+1, \frac{\beta}{2}+1\right)}, \qquad K_1^{\perp} = \frac{\kappa_{\perp}}{\mathrm{B}\left(\frac{\beta}{2}+1, \frac{\beta}{2}+1\right)} \qquad \text{with} \qquad K_1^{\perp} = \frac{2}{1+\beta}K_1^{=}, \tag{50}$$

and the functions $\mu_1^{\perp}(s) = K_1^{\perp}h(s)$, and $\mu_1^{=}(s) = K_1^{=}h(s)$.

Proposition 20. *There exists $\varepsilon_{=} > 0$ such that*

$$\sup_{s\in[0,1]} \left|n^{-\beta}\mathbf{E}\left[C_n^{=}(s)\right] - \mu_1^{=}(s)\right| = O(n^{-\varepsilon_{=}}),$$

and the analogous result holds true for $\mathbf{E}[C_n^{\perp}(s)]$.

We proceed as in Section 5 by considering the continuous-time process $P_t^{=}(s)$. Since we have already proved an analogous result for the case of quadtree, we give a brief sketch that focuses on the few locations where the arguments have to be modified.

Sketch of proof. The first step is to prove point-wise convergence which is done as Curien and Joseph [6]. By Lemma 19, using a Poisson(t) number of points, we have

$$\frac{1}{5}\mathbf{E}[P_t(s)] \leq \mathbf{E}[P_t^{=}(s)] \leq 2\mathbf{E}[P_t(s)]. \tag{51}$$

Let $\tau_1^{=}$ be the arrival time of the first point which yields a partitioning line that intersects the query line $\{x = s\}$, and let $Q_1^{=} = Q_1^{=}(s)$ be the lower of the two rectangles created by this cut (for the expected value we are about to compute, they both look the same). Let $\xi_1^{=} := \xi_1^{=}(s)$ be the relative position of the query line s within the rectangle $Q_1^{=}$ and $M_1^{=} = \text{Leb}(Q_1^{=})$. Then, denoting τ the arrival time of the first point in the process, we have

$$\mathbf{E}[P_t^{=}(s)] = \mathbf{P}\left(t \geq \tau\right) + \mathbf{P}\left(t \geq \tau_1^{=}\right) + 2\mathbf{E}[\tilde{P}^{=}_{M_1^{=}t-\tau_1^{=}}(\xi_1^{=})],$$

where $(\tilde{P}^{=}(t))_{t\geq 0}$ denotes an independent copy of $(P^{=}(t))_{t\geq 0}$ and $\tilde{P}^{=}(t) = 0$ for $t < 0$. Similarly, let $\tau_k^{=}$ be the arrival time of the first point which cuts $Q_{k-1}^{=}$ perpendicularly to the query line. Let $Q_k^{=}$ be the lower of the two rectangles created by this cut, and let $\xi_k^{=}$ be the position of the query line s relative to the rectangle $Q_k^{=}$. With this notation and $M_k^{=} = \text{Leb}(Q_k^{=})$, we have

$$\mathbf{E}[P_t^{=}(s)] = g_k^{=}(t) + 2^k\mathbf{E}[\tilde{P}^{=}_{M_k^{=}t-\tau_k^{=}}(\xi_k^{=})],$$

where $0 \leq g_k^{=}(t) \leq 2^{k+1}$.

We need to modify the inter-arrival times $\zeta'^{=}_k = \tau_k^{=} - \tau_{k-1}^{=}$. We can split $\zeta'^{=}_k$ in the time it takes for the first vertical point to fall in $Q_{k-1}^{=}$ which we denote by $\zeta'^{=,1}_k$ and the remaining time by $\zeta'^{=,2}_k$. Letting $M_k^{=} = \text{Leb}(Q_k^{=})$, the normalized versions of the inter-arrival times with unit mean are

$$\begin{aligned}
\zeta_k^{=,1} &= \zeta'^{=,1}_k \cdot M_{k-1}^{=},\\
\zeta_k^{=,2} &= \left(\frac{\xi_k^{=}}{\xi_{k-1}^{=}}\mathbf{1}_{\{\xi_k^{=}<\xi_{k-1}^{=}\}} + \frac{\xi_{k-1}^{=}}{\xi_k^{=}}\mathbf{1}_{\{\xi_k^{=}\geq\xi_{k-1}^{=}\}}\right)\zeta'^{=,2}_k \cdot M_{k-1}^{=} \geq \zeta'^{=,2}_k \cdot M_{k-1}^{=}.
\end{aligned}$$

Write $\mathcal{M}_k = M_k/M_{k-1}$. Observe that, given $\mathcal{M}_0^{=}, \ldots, \mathcal{M}_k^{=}$, the random variable $F_k^{=} = M_k^{=} \cdot \tau_k^{=}$ is not independent of $(\xi_\ell)_{0\leq\ell\leq k}$, a property which is used in [6] and in the proof of Lemma 15 in the present paper. However we can use the trivial lower bound $0 \leq F_k$ and the upper bound obtained by bounding $\zeta'^{=,2}_k$ from above by $\zeta_k^{=,2}/M_{k-1}^{=}$. Then, using almost sure monotonicity of $P_t(s)$ (in t) and (51) to transform bounds for the mean in the quadtree to bounds in the 2-d tree (and vice versa), it is easy to see that the techniques of Section 4 in [6] work equally well in this case. The limit $\mu_1^{=}(s)$ is identified as in Section 5 of [6] since both limits satisfy the same fixed-point equation.

The generalization to uniform convergence with polynomial rate can be worked out as in Section 5 (of the present document) using the modifications we have described above. The constants appearing in the course of Section 5 need to be modified, but $\varepsilon_{=}$ may be chosen to equal the value of ε in Proposition 13. The de-Poissonization is routine and we omit the details.

Finally, we indicate how to proceed with $\mathbf{E}[C_n^{\perp}(s)]$. The arguments above can be used to treat prove uniform convergence of $n^{-\beta}\mathbf{E}[C_n^{\perp}(s)]$ on $[0, 1]$; we present a direct approach relying on (39). We have

$$\begin{aligned}
n^{-\beta}\mathbf{E}[C_n^{\perp}(s)] &= n^{-\beta} + 2n^{-\beta}\mathbf{E}\left[C^{=}_{S_n}(s)\right]\\
&= n^{-\beta} + 2\int_0^1 \sum_{k=0}^{n-1}\left(\mu_1^{=}(s) + O(k^{-\varepsilon_{=}})\right)\frac{k^\beta}{n^\beta}\mathbf{P}\left(\text{Bin}(n-1, v) = k\right)dv\\
&= n^{-\beta} + 2\mu_1^{=}(s)\cdot\frac{\mathbf{E}\left[\text{Bin}(n-1, V)^\beta\right]}{n^\beta} + O(n^{-\beta}\mathbf{E}[\text{Bin}(n-1, V)^{\beta-\varepsilon_{=}}])\\
&= \mu_1^{\perp}(s) + O(n^{-\varepsilon_{=}}),
\end{aligned}$$

uniformly in $s \in [0, 1]$ using Minkowski's inequality, the concentration for binomial in (21), and (50) for the first term and Jensen's inequality for the second. □

7.3 The limiting behaviour in 2-d trees

We are finally ready to state the version of our main result for 2-d trees. It is proved along the same lines we used for the case of quadtrees, and we omit the details.

Theorem 21. *With the processes $Z^=$ and $Z^\perp$ of Proposition 18 we have*

$$\left(\frac{C_n^=(s)}{K_1^= n^\beta}\right)_{s\in[0,1]} \to (Z^=(s))_{s\in[0,1]}, \qquad \left(\frac{C_n^\perp(s)}{K_1^\perp n^\beta}\right)_{s\in[0,1]} \to \left(Z^\perp(s)\right)_{s\in[0,1]},$$

in distribution in $\mathcal{D}[0,1]$ endowed with the Skorokhod topology. Here $K_1^=$ and $K_1^\perp$ are defined in (50). For $s \in [0,1]$

$$n^{-\beta}\mathbf{E}[C_n^=(s)] \to K_1^= h(s), \quad n^{-2\beta}\mathbf{Var}\,(C_n^=(s)) \to (K_1^=)^2 K_2 h(s)^2,$$

and

$$n^{-\beta}\mathbf{E}[C_n^\perp(s)] \to K_1^\perp h(s), \quad n^{-2\beta}\mathbf{Var}\left(C_n^\perp(s)\right) \to (K_1^\perp)^2 K_2^\perp h(s)^2,$$

where K_2 is given in (8) and $K_2^\perp$ in (48).

If ξ is uniformly distributed on $[0,1]$, independent of $(C_n^=)_{n\ge 0}, (C_n^\perp)_{n\ge 0}$ and $Z^=, Z^\perp$, then

$$\frac{C_n^=(\xi)}{K_1^= n^\beta} \xrightarrow{d} Z^=(\xi), \qquad \frac{C_n^\perp(\xi)}{K_1^\perp n^\beta} \xrightarrow{d} Z^\perp(\xi),$$

with convergence of the first two moments in both cases. In particular

$$\mathbf{Var}\,(C_n^=(\xi)) \sim K_4^= n^{2\beta}, \qquad \mathbf{Var}\left(C_n^\perp(\xi)\right) \sim K_4^\perp n^{2\beta},$$

where $K_4^= = (K_1^=)^2 K_3 \approx 0.69848$, $K_4^\perp = (K_1^\perp)^2 K_3^V \approx 0.77754$, with $K_3 = \mathbf{Var}\,(Z(\xi))$ in (7) and $K_3^\perp$ in (49).

Note that since $Z^=(s)$ equals $Z(s)$ in distribution for fixed $s \in [0,1]$ we can characterize $Z^=(s)$ as in (9). (47) together with Proposition 18 implies that for fixed $s \in [0,1]$

$$Z^\perp(s) \stackrel{d}{=} \Xi^\perp \cdot h(s) \qquad \text{with} \qquad \Xi^\perp = \frac{\beta+1}{2}\left(V^\beta \Xi + (1-V)^\beta \Xi'\right),$$

where Ξ' is an independent copy of Ξ, Ξ being defined in Theorem 5 and V is independent of (Ξ, Ξ'). In particular, we have

$$\mathbf{E}\left[(\Xi^\perp)^m\right] = \left(\frac{\beta+1}{2}\right)^m \sum_{\ell=0}^m \binom{m}{\ell} \mathrm{B}(\beta\ell+1, \beta(m-\ell)+1) c_\ell c_{m-\ell},$$

for $m \ge 2$ where $c_m = \mathbf{E}[\Xi^m]$ satisfies recursion (11) and $c_0 = c_1 = 1$.

Also, as in the quadtree case, it is possible to give convergence of mixed moments of arbitrary order, compare Corollary 6, and distributional and moment convergence of the suprema of the processes after rescaling as in Theorem 4.

References

[1] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communication of the ACM*, 18:509–517, 1975.

[2] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. Wiley, second edition, 1999.

[3] N. Broutin, R. Neininger, and H. Sulzbach. Partial match queries in random quadtrees. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1056–1065, 2012.

[4] H.-H. Chern and H.-K. Hwang. Partial match queries in random quadtrees. *SIAM Journal on Computing*, 32:904–915, 2003.

[5] H.-H. Chern and H.-K. Hwang. Partial match queries in random k-d trees. *SIAM Journal on Computing*, 35:1440–1466, 2006.

[6] N. Curien and A. Joseph. Partial match queries in two-dimensional quadtrees: A probabilistic approach. *Advances in Applied Probability*, 43:178–194, 2011.

[7] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24:277–298, 1987.

[8] L. Devroye and L. Laforest. An analysis of d-dimensional quad trees. *SIAM Journal on Computing*, 19:821–832, 1990.

[9] M. Drmota, S. Janson, and R. Neininger. A functional limit theorem for the profile of search trees. *The Annals of Applied Probability*, 18(1):288–333, 2008.

[10] A. Duch and C. Martínez. On the average performance of orthogonal range search in multidimensional data structures. *Journal of Algorithms*, 44(1):226–245, 2002.

[11] A. Duch, V. Estivill-Castro, and C. Martínez. Randomized k-dimensional binary search trees. In K.-Y. Chwa and O. Ibarra, editors, *Proc. of the 9th International Symposium on Algorithms and Computation (ISAAC'98)*, volume 1533 of *Lecture Notes in Computer Science*, pages 199–208. Springer Verlag, 1998.

[12] A. Duch, R. Jiménez, and C. Martínez. Rank selection in multidimensional data. In A. López-Ortiz, editor, *Proceedings of LATIN*, volume 6034 of *Lecture Notes in Computer Science*, pages 674–685, Berlin, 2010. Springer.

[13] K. Eickmeyer and L. Rüschendorf. A limit theorem for recursively defined processes in L^p. *Statistiscs & Decisions*, 25(3):217–235, 2007.

[14] W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. Wiley, New York, 3rd edition, 1971.

[15] R. A. Finkel and J. L. Bentley. Quad trees, a data structure for retrieval on composite keys. *Acta Informatica*, 4:1–19, 1974.

[16] P. Flajolet and T. Lafforgue. Search costs in quadtrees and singularity perturbation asymptotics. *Discrete and Computational Geometry*, 12:151–175, 1994.

[17] P. Flajolet and C. Puech. Partial match retrieval of multidimensional data. *Jounal of the ACM*, 33(2): 371–407, 1986.

[18] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, UK, 2009.

[19] P. Flajolet, G. H. Gonnet, C. Puech, and J. M. Robson. Analytic variations on quadtrees. *Algorithmica*, 10:473–500, 1993.

[20] P. Flajolet, G. Labelle, L. Laforest, and B. Salvy. Hypergeometrics and the cost structure of quadtrees. *Random Structures and Algorithms*, 7:117–144, 1995.

[21] R. Grübel. On the silhouette of binary search trees. *The Annals of Applied Probability*, 19:1781–1802, 2009.

[22] K. Ho-Le. Finite element mesh generation methods: a review and classification. *Computer-Aided Design*, 20:27–38, 1988.

[23] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[24] D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, 2d edition, 1998.

[25] H. Mahmoud. *Evolution of Random Search Trees*. Wiley, New York, 1992.

[26] C. Martínez, A. Panholzer, and H. Prodinger. Partial match in relaxed multidimensional search trees. *Algorithmica*, 29(1–2):181–204, 2001.

[27] R. Neininger. Asymptotic distributions for partial match queries in K-d trees. *Random Structures and Algorithms*, 17:403–427, 2000.

[28] R. Neininger. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, 19:498–524, 2001.

[29] R. Neininger and L. Rüschendorf. Limit laws for partial match queries in quadtrees. *The Annals of Applied Probability*, 11:452–469, 2001.

[30] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, 14:378–418, 2004.

[31] R. Neininger and L. Rüschendorf. On the contraction method with degenerate limit equation. *The Annals of Probabability*, 32:2838–2856, 2004.

[32] R. Neininger and H. Sulzbach. On a functional contraction method. preprint available `http://arxiv.org/`.

[33] S. Rachev and L. Rüschendorf. Probability metrics and recursive algorithms. *Advances in Applied Probability*, 27:770–799, 1995.

[34] R. Rivest. Partial-match retrieval algorithms. *SIAM Journal on Computing*, 5(19–50), 1976.

[35] U. Rösler. A limit theorem for "quicksort". *RAIRO Informatique théorique et Applications*, 25:85–100, 1991.

[36] U. Rösler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 37: 195–214, 1992.

[37] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29:238–261, 2001.

[38] H. Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, MA, 1990.

[39] H. Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*. Addison-Wesley, Reading, MA, 1990.

[40] H. Samet. *Foundations of multidimensional and metric data structures*. Morgan Kaufmann, 2006.

[41] M. Yerry and M. Shephard. A modified quadtree approach to finite element mesh generation. *IEEE Computer Graphics and Applications*, 3:39–46, 1983.

A About the geometry of random quadtrees

Lemma 22. *Let W_n denote the maximum width of a cell at level n in the construction of Z_n and $c < 1$. Then,*

$$\mathbf{P}\left(W_n \geq c^n\right) \leq (4e\log(1/c))^n.$$

Proof. Let U_i, $i \geq 1$ be a family of i.i.d. $[0,1]$-uniform random variables and E_i, $i \geq 1$, be a family of i.i.d. exponential(1) random variables. Then, the union bound and a large deviations argument yields

$$\begin{aligned}
\mathbf{P}\left(W_n \geq c^n\right) &\leq 4^n \cdot \mathbf{P}\left(U_1 \cdot U_2 \cdots U_n \geq c^n\right) \\
&= 4^n \cdot \mathbf{P}\left(\sum_{i=1}^n E_i \leq n\log(1/c)\right) \\
&\leq 4^n \exp(-n(\log(1/c) - 1 - \log\log(1/c))) \\
&\leq (4e\log(1/c))^n,
\end{aligned}$$

as desired. □

Lemma 23. *Let F_k be the fill-up level of a random quadtree of size k. Then, for every integer number $x > 22$ there exists an integer $n_0(x)$ with*

$$\mathbf{P}\left(F_{x^n} < n\right) \leq 4^{n+1}x^{-n/100}, \qquad n \geq n_0(x).$$

Proof. We consider the 4^n possible nodes in level n. By symmetry each of them is occupied by a key with the same probability. Looking at a specific one, e.g. the leftmost, size of the corresponding subtree is stochastically bounded by $\text{Bin}(x^n; U_1V_1 \cdots U_nV_n) - n$ where $\{U_i, i \geq 1\}$ and $\{V_i, i \geq 1\}$ are independent families of i.i.d. $[0,1]$-uniform random variables. Then by the union bound applied to the 4^n cells at level n, using Chernoff's inequality, we have

$$\begin{aligned}
\mathbf{P}\left(F_{x^n} < n\right) &\leq 4^n \cdot \mathbf{P}\left(\text{Bin}\left(x^n; U_1V_1 \cdots U_nV_n\right) \leq n\right) \\
&\leq 4^n \cdot \exp\left(-(1-n2^{-n})^2 2^{n+1}\right) + 4^n \mathbf{P}\left(U_1V_1 \cdots U_nV_n \leq \left(\frac{2}{x}\right)^n\right). \qquad (52)
\end{aligned}$$

However, using once again the large deviations principle for sums of i.i.d. exponential random variables $E_i, i \geq 1$,

$$\begin{aligned}\mathbf{P}\left(U_1V_1\dots U_nV_n \leq (2/x)^n\right) &= \mathbf{P}\left(\sum_{i=1}^{2n} E_i \geq n\log(x/2)\right)\\ &\leq \exp\left(-2n\left(\frac{\log(x/2)}{2} - 1 - \log\frac{\log(x/2)}{2}\right)\right)\\ &\leq x^{-n/100}, \end{aligned} \tag{53}$$

for all $x > 22$ since then $\frac{e^2}{2}\log^2(x/2) \leq x^{99/100}$. Putting (52) and (53), we obtain

$$\mathbf{P}\left(F_{x^n} < n\right) \leq 4^n \exp(-2^{n-1}) + 4^n \cdot x^{-n/100} \leq 4^{n+1}x^{-n/100},$$

for $x > 22$ and n large enough. □

Lemma 24. *There exists $0 < \gamma_0 < 1$ such that any positive real number $\gamma < \gamma_0$, there exists an integer $n_1(\gamma)$ with*

$$\mathbf{P}\left(L_n < \gamma^n\right) \leq 6^{n+1}\gamma^{n/201}, \qquad n \geq n_1(\gamma).$$

Proof. The joint distribution of the x-coordinates of the vertical lines in the tree developed up to level n is complex. In particular, it is *not* that of independent uniform points on $[0,1]$. However, we can use a simple coupling with a family of i.i.d. random points on $[0,1]^2$ that yields a good enough lower bound on L_n.

Let $\xi_i = (U_i, V_i)$, $i \geq 1$ be i.i.d. uniform random points on $[0,1]^2$. Let T_k be the quadtree obtained by inserting the random points ξ_i, $1 \leq i \leq k$, in this order. Write D_i for the depth at which the point ξ_i is inserted ; so for instance $D_1 = 0$. Let K_n be the first k for which the tree T_k is complete up to level n; we mean here that T_k should have 4^n cells at level n, so it should have 4^{n-1} nodes at level $n-1$. Then, by definition $\{\xi_i : i \geq 1, D_i < n\}$ has the distribution of the set of points used to construct the process Z_n. Obviously, $\{\xi_i : i \geq 1, D_i < n\} \subseteq \{\xi_i : 1 \leq i \leq K_n\}$ and for any natural number $x > 0$,

$$\begin{aligned}\mathbf{P}\left(L_n < \gamma^n\right) &\leq \mathbf{P}\left(\exists i,j \leq K_n : i \neq j, |U_i - U_j| < \gamma^n\right)\\ &\leq \mathbf{P}\left(\exists i,j \leq x^n : i \neq j, |U_i - U_j| < \gamma^n\right) + \mathbf{P}\left(K_n > x^n\right)\\ &\leq x^{2n}\cdot 2\gamma^n + \mathbf{P}\left(K_n > x^n\right),\end{aligned}$$

by the union bound. The random variable K_n is related to the fill-up level of a random quadtree, which has been studied by [7] (see also [8]). We could not find a reference giving a precise tail bound, so we proved one here in Lemma 23. We obtain

$$\mathbf{P}\left(K_n > x^n\right) = \mathbf{P}\left(F_{x^n} < n\right) \leq 4(4x^{-1/100})^n,$$

as long as $x \geq 22$ and $n \geq n_0(x)$ (the condition for the bound in Lemma 23 to hold). It follows readily that

$$\begin{aligned}\mathbf{P}\left(L_n < \gamma^n\right) &\leq 2(x^2\gamma)^n + 4(4x^{-1/100})^n\\ &\leq 6^{n+1}\gamma^{n/201},\end{aligned}$$

upon choosing $x = \lceil 4^{100/201}\gamma^{-100/201}\rceil$ (that is $x^2\gamma \approx 4x^{-1/100}$) and $\gamma < 4\cdot 22^{-2.01}$ which implies $x > 22$. This completes the proof. □

B Complements to the proof of Propopsition 12

B.1 Behavior away from the edge: proof of Lemma 15

The core of the work is to bound the second term in (33) involving $s \in (\delta, 1/2]$. We prove that $\mathbf{E}[P_t(s)]$ is uniformly Cauchy on $(\delta, 1/2]$ by tightening some of the arguments in [6]. We could start from (14) there, but we feel that the reader would follow more easily if we re-explain the approach. Observe that most of the quantities defined in the remaining of the section will depend on s which we will neglect in the notation for the sake of readability.

The first step is to unfold k levels of the fundamental recurrence (13) in the Poisson case. Let τ_1 be the arrival time of the first point in the Poisson process and $Q_1 = Q_1(s)$ be the lower of the two rectangles that intersect the line $\{x = s\}$ after inserting the first point. Inductively let $\tau_k = \tau_k(s)$ be the arrival time of the first point of the process in the region Q_{k-1} and Q_k be the lower of the two rectangles that hit the line $\{x = s\}$ at time τ_k. For convenience, set $Q_0 = [0,1]^2$. Finally, let $\tilde{P}_t$ be an independent copy of the process P_t (set $\tilde{P}_t \equiv 0$ for $t < 0$). At level one, using the horizontal symmetry, we have

$$\mathbf{E}[P_t(s)] = \mathbf{P}(t \geq \tau_1) + 2\mathbf{E}[\tilde{P}_{\mathrm{Leb}(Q_1)(t-\tau_1)}(\xi_1)],$$

where $\xi_1 = \xi_1(s) \in [0,1]$ denotes the location of the line $\{x = s\}$ relative to the region Q_1. If the interval $[\ell_1, r_1]$ denotes the projection of Q_1 on the first axis, we have

$$\xi_1(s) = \frac{s - \ell_1}{r_1 - \ell_1}.$$

Write $\xi_k = \xi_k(s) \in [0,1]$ for the location of the line $\{x = s\}$ relatively to the region Q_k, and $M_k = \mathrm{Leb}(Q_k)$. Then, unfolding up to level k, we obtain

$$\mathbf{E}[P_t(s)] = g_k(t) + 2^k\mathbf{E}[\tilde{P}_{M_k(t-\tau_k)}(\xi_k)], \tag{54}$$

where $0 \leq g_k(t) \leq 2^k - 1$. Next, we introduce the inter-arrival times $\zeta'_k = \tau_k - \tau_{k-1}$ with $\zeta'_0 := 0$ and their normalized versions $\zeta_k = \zeta'_k M_{k-1}$ (again $\zeta_0 := 0$). Defining $F_k = M_k \tau_k$, we can rewrite (54) as

$$\mathbf{E}[P_t(s)] = g_k(t) + 2^k\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)]. \tag{55}$$

Note that $(\zeta_k)_{k\geq 1}$ are i.i.d. exponential random variables with unit mean, also independent of $(\xi_k, Q_k)_{k\geq 1}$.

Before going any further, note that, as we have already seen in Section 4, the region Q_k, is not distributed like a typical rectangle at level k; in particular $\mathrm{Leb}(Q_k)$ is not distributed as $X_1Y_1 \cdots X_kY_k$, for independent $[0,1]$-uniform random variables X_i, Y_i, $i \geq 1$. Intuitively, Q_k should be stochastically larger than a typical cell, since it is conditioned to intersect the line $\{x = s\}$. This is verified by the following lemma.

Lemma 25. *For any $s \in (0,1)$, any integer $k \geq 0$, and $1 \leq i \leq 2^k$, we have*

$$\mathrm{Leb}(Q_k) = M_k \geq_{st} X_1Y_1 \cdots X_kY_k,$$

where X_i, Y_i, $i \geq 1$ are independent random variables uniform on $[0,1]$.

Proof. Consider one split, at a point (X, Y) uniform inside the unit square. The split creates four new boxes, two of them being hit by s. Let L be the length these two cells. Their height is either Y or $(1 - Y)$, which are both uniform. So it suffices to prove that $L \geq_{\mathrm{st}} X$. By symmetry, it suffices to consider $s \leq 1/2$. We have,

$$L = \mathbf{1}_{\{s \leq X\}} X + \mathbf{1}_{\{s > X\}}(1 - X).$$

Write $F_L(y) = \mathbf{P}(L \leq y)$ and $F_X(y) = \mathbf{P}(X \leq y) = y$. It is then easy to see that

$$F_L(y) = \mathbf{P}(L \leq y) = \begin{cases} 0, & y \leq s \\ y - s, & s \leq y \leq 1 - s \\ 2y - 1, & y \geq 1 - s. \end{cases}$$

Hence, for all $s \in (0, 1/2)$ and all $y \in (0,1)$ we have $F_L(y) \leq y = F_X(y)$. The result follows. □

The second term will be treated using results for the case $s = \xi$, for a uniform random variable ξ independent of everything else. Curien and Joseph [6] found a very clever way to circumvent the problem that for any $k \geq 1$, the random variable ξ_k is not uniformly distributed on $[0,1]$. In their Proposition 4.1 they introduce a version of the homogeneous Markov chain $(\xi_k, \mathcal{M}_k)_{k\geq 1}$ where $\mathcal{M}_k := M_k/M_{k-1}$ together with a random time T such that for any $k \in \mathbb{N}$, conditionally on $\{T \leq k\}$, the random variable ξ_k is uniformly distributed on $[0,1]$, independent of $(\mathcal{M}_1, \dots, \mathcal{M}_k, T)$. Choosing these random variables independent of the process $\tilde{P}_t$ we will use them in the following without changing the notation (F_k can be constructed using $(\mathcal{M}_\ell)_{1\leq \ell \leq k}$ and an additional set of i.i.d. exponential random variables with mean one).

The details of the definition of T are not important for us. The only crucial thing is that T has exponential tails. Indeed, we have [p.15 of 6]

$$\mathbf{E}[1.15^T] \le C_4(s \wedge (1-s))^{-1/2} \le C_4\delta^{-1/2}, \tag{56}$$

for some constant C_4 in the present case, $\delta < s \le 1/2$.

Then, using (55) and the triangle inequality, we obtain for any t and r such that $r \ge t$,

$$\begin{aligned} \left|t^{-\beta}\mathbf{E}[P_t(s)] - r^{-\beta}\mathbf{E}[P_r(s)]\right| \le& 2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)]\right| + 2^{k+1}r^{-\beta} \\ \le& 2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T\le k\}}] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)\mathbf{1}_{\{T\le k\}}]\right| \\ &+ 2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}]\right| \\ &+ 2^{k+1}r^{-\beta}. \end{aligned} \tag{57}$$

To complete the proof of Lemma 15, we now devise explicit bounds for the two main terms in (57) when we can ensure that coupling occured by level k (i.e., $T \le k$) or not.

i. No coupling by level k, $T > k$. In this case, we bound the terms roughly. We obtain

$$\begin{aligned} &2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}]\right| \\ &\le 2^{k+1}\sup_{u\ge t} u^{-\beta}\mathbf{E}[\tilde{P}_{M_k u - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}]. \end{aligned}$$

One then essentially uses the uniform bound $\sup_s \sup_u u^{-\beta}\mathbf{E}[P_u(s)] \le C_5$ (see (10) in [6]) Hölder's and Markov's inequalities to leverage a bound that makes profit of the exponential tails of T. The details are found in [6], p. 16. For any $u > 0$ and $s \in (\delta, 1/2]$, one has

$$\begin{aligned} u^{-\beta}2^k\mathbf{E}[\tilde{P}_{M_k u - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}] &\le C_5 2^k s^{-1/p}\left(\frac{2}{(\beta p+1)(\beta p+2)}\right)^{(k-1)/p}\left(\frac{\mathbf{E}[1.15^T]}{1.15^k}\right)^{1-1/p} \\ &\le C_4C_5\delta^{-1/2-1/(2p)}\left(2\left\{\frac{2}{(\beta p+1)(\beta p+2)}\right\}^{1/p} 1.15^{1/p-1}\right)^k, \end{aligned}$$

by the upper bound in (56). Choosing p close enough to one that the term in the brackets above is strictly less than one, we obtain for any $s \in (\delta, 1/2]$ and real numbers $t, r > 0$,

$$\begin{aligned} 2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)\mathbf{1}_{\{T> k\}}]\right| &\le 2C_4C_5\delta^{-1/2-1/(2p)}(1-\gamma)^k \\ &\le C_1\delta^{-1}(1-\gamma)^k, \end{aligned} \tag{58}$$

where C_1 denotes a constant and $\gamma > 0$ (and $p > 1$ is now fixed).

ii. Coupling has occurred before level k, $T \le k$. In this case, we need to be a little more careful and match some terms. In what follows, we write $x_+ = x \vee 0$. We start with

$$t^{-\beta}2^k\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T\le k\}}] = 2^k\mathbf{E}[\mathbf{1}_{\{T\le k\}}(M_k - t^{-1}F_k)_+^\beta\theta(M_k t - F_k)],$$

where $\theta(x) = x_+^{-\beta}\mathbf{E}[P_x(X)]$ with X a $[0,1]$-uniform random variable independent of everything else. The estimate in (2) is easily transferred to the Poissonized version, and we have $\theta(x) = \kappa + O(x^{-\eta})$ for any $0 < \eta < \beta$. Therefore

$$\begin{aligned} &2^k\left|t^{-\beta}\mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k)\mathbf{1}_{\{T\le k\}}] - r^{-\beta}\mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k)\mathbf{1}_{\{T\le k\}}]\right| \\ &\le 2^k\left|\mathbf{E}[\mathbf{1}_{\{T\le k\}}(M_k - t^{-1}F_k)_+^\beta\theta(M_k t - F_k)] - \mathbf{E}[\mathbf{1}_{\{T\le k\}}(M_k - r^{-1}F_k)_+^\beta\theta(M_k r - F_k)]\right| \\ &\le 2^k\mathbf{E}\left[\left|(M_k - t^{-1}F_k)_+^\beta\theta(M_k t - F_k) - (M_k - r^{-1}F_k)_+^\beta\theta(M_k r - F_k)\right|\right]. \end{aligned} \tag{59}$$

Fix $\eta < \beta$. For $x > 0$, we have, as $x \to \infty$

$$\begin{aligned} (M_k - x^{-1}F_k)_+^\beta \cdot \theta(M_k x - F_k) &= M_k^\beta(1 - O(x^{-1}F_kM_k^{-1}))(\kappa + O(M_k^{-\eta}x^{-\eta})) \\ &= \kappa M_k^\beta + O(F_kM_k^{\beta-1}x^{-1}) + O(M_k^{\beta-\eta}x^{-\eta}) + O(F_kM_k^{\beta-1-\eta}x^{-1-\eta}) \\ &= \kappa M_k^\beta + O(F_kM_k^{\beta-1}x^{-1}) + O(x^{-\eta}) + O(F_kM_k^{\beta-1-\eta}x^{-1-\eta}), \end{aligned}$$

since $M_k \in (0,1)$ and $\eta < \beta$, the $O(\cdot)$ terms being deterministic and uniform in $s \in [0,1]$. Going back to (59), the terms κM_k^β coming from the two terms with t and r cancel out, and there exist constants C_7, C_8 such that, for all t, r large enough such that moreover $t \le r$, we have

$$\begin{aligned}
&2^k \big| t^{-\beta} \mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k) \mathbf{1}_{\{T \le k\}}] - r^{-\beta} \mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k) \mathbf{1}_{\{T \le k\}}] \big| \\
&\le C_7 2^k \left(t^{-1} \mathbf{E}[F_k M_k^{\beta-1}] + t^{-\eta} + t^{-1-\eta} \mathbf{E}[F_k M_k^{\beta-1-\eta}] \right) \\
&\le C_8 2^k t^{-\eta} \mathbf{E}[F_k M_k^{\beta-1-\eta}].
\end{aligned}$$

Since it will be necessary to choose k tending to infinity with r to control the term in (58), it remains to estimate $\mathbf{E}[F_k M_k^{\beta-1-\eta}]$. By definition of $F_k = M_k \tau_k$, one easily verifies that $F_k \le \sum_{i=1}^k \zeta_k$, where the normalized inter-arrival times ζ_i were defined right after (54). Since $M_i \le 1$ for every i, we have

$$\begin{aligned}
\mathbf{E}[F_k M_k^{\beta-1-\eta}] &\le k \mathbf{E}[M_k^{\beta-1-\eta}] \\
&\le k \mathbf{E}[X^{\beta-1-\eta}]^{2k} = k(\beta-\eta)^{-2k},
\end{aligned}$$

by the lower bound on M_k in Lemma 25, X denoting a uniform on $[0,1]$. We finally obtain

$$2^k \big| t^{-\beta} \mathbf{E}[\tilde{P}_{M_k t - F_k}(\xi_k) \mathbf{1}_{\{T \le k\}}] - r^{-\beta} \mathbf{E}[\tilde{P}_{M_k r - F_k}(\xi_k) \mathbf{1}_{\{T \le k\}}] \big| \le C_8 k t^{-\eta} 2^k (\beta-\eta)^{-2k}. \tag{60}$$

Putting (58) and (60) together with (57) yields, for any $t, r > 0$ such that $t \le r$

$$\begin{aligned}
\big| t^{-\beta} \mathbf{E}[P_t(s)] - r^{-\beta} \mathbf{E}[P_r(s)] \big| &\le C_1 \delta^{-1} (1-\gamma)^k + C_8 k 2^k (\beta-\eta)^{-2k} t^{-\eta} + 2^{k+1} t^{-\beta} \\
&\le C_1 \delta^{-1} (1-\gamma)^k + C_2 k 2^k (\beta-\eta)^{-2k} t^{-\eta}.
\end{aligned}$$

for some constant C_2. The statement in Lemma 15 follows readily from the triangle inequality.