SIAM J. COMPUT.
Vol. 35, No. 6, pp. 1440–1466

PARTIAL MATCH QUERIES IN RANDOM k-d TREES*

HUA-HUAI CHERN† AND HSIEN-KUEI HWANG‡

Abstract. We solve the open problem of characterizing the leading constant in the asymptotic approximation to the expected cost used for random partial match queries in random k-d trees. Our approach is new and of some generality; in particular, it is applicable to many problems involving differential equations (or difference equations) with polynomial coefficients.

Key words. k-d trees, partial match queries, differential equations, average-case analysis of algorithms, method of linear operators, asymptotic analysis

AMS subject classifications. 68W40, 68P05, 68P10, 68U05

DOI. 10.1137/S0097539703437491

1. Introduction.

Multidimensional binary search trees (abbreviated as k-d trees or simply kd trees, k being the dimensionality) were first proposed by Bentley [2] and represent extremely useful data structures for problems in diverse fields, especially those having to do with range queries; nearest neighbor search; or partial, exact, or approximate match queries. For example, they are useful in statistical learning, databases, data-mining, computer graphics, robotics, medical imaging, neural networks, multimedia, statistical computing, computer-aided design, astronomy, pattern recognition, geographic information systems, music information retrieval, computational biology, etc. For more information, see Bentley [3], Bentley and Friedman [4], Bertino et al. [5], Friedman, Bentley, and Finkel [19], Gray and Moore [21], Grother, Candela, and Blue [22], Nichol et al. [28], Omohundro [29], Preparata and Shamos [31], Reiss, Aucouturier, and Sandler [32], Samet [33, 34], and Schwarzer and Lotan [35]. According to the statistics collected in the Stony Brook Algorithm Repository, k-d trees were among the most popular algorithmic problems; see Skiena's account in [37].

Despite the usefulness and diversity of k-d trees, their precise probabilistic analysis appears only sporadically in the literature (although many properties can be derived from those for random binary search trees); see [6, 11, 12, 13, 19, 20, 25, 26, 27]. We are concerned in this paper with the expected cost used by random partial match queries in random k-d trees. The growth order of this cost has been known since Flajolet and Puech's work [17], but the characterization of the leading constant remains a very challenging problem. We propose a new approach to filling this gap.

The prototypes of k-d trees are binary search trees when there is a total ordering for the input keys or when $k = 1$. A binary search tree $\mathcal{B}$ is a binary tree constructed from a given sequence of keys as follows. If the tree size is $n = 0$, then $\mathcal{B}$ is empty. If $n \geq 1$, then the first key is placed at the root. The remaining keys are compared successively to the root key, which may be called a "discriminator," and are directed to the left (or right) branch if they are smaller (or larger); keys directed to the same

*Received by the editors November 10, 2003; accepted for publication (in revised form) December 8, 2005; published electronically April 7, 2006.
http://www.siam.org/journals/sicomp/35-6/43749.html
†Department of Computer Science, National Taiwan Ocean University, 2 Pei-Ning Road, Keelung 20224, Taiwan (felix@cs.ntou.edu.tw).
‡Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan (hkhwang@stat.sinica.edu.tw).

branch are constructed recursively as a binary search tree. By construction, a query operation such as "$x \in \mathcal{B}$?" can be easily carried out in binary search trees; hence the name "search tree."

When given k-dimensional points or keys, $k \geq 2$, the simple idea of k-d trees is to use each coordinate cyclically, say, in the order from the first coordinate to the last, as the "discriminator" to direct points falling in the same subtree as in binary search trees (but using the $(\ell \bmod k) + 1$st coordinate for a node at a distance of ℓ from the root); see Figure 1.1 for a plot of a simple 2-d tree.

In addition to exact match searches, as in binary search trees, k-d trees can also be used for *partial match queries* when some coordinates are either unspecified, don't-cares, or wild cards. Thus the range search is continued in both subtrees when the unspecified coordinate is used as a discriminator in the k-d tree and in either the left or the right subtree (but not both) otherwise; see Bentley [2] for details. Two simple instances of a range search in 2-d trees are shown in Figure 1.2.

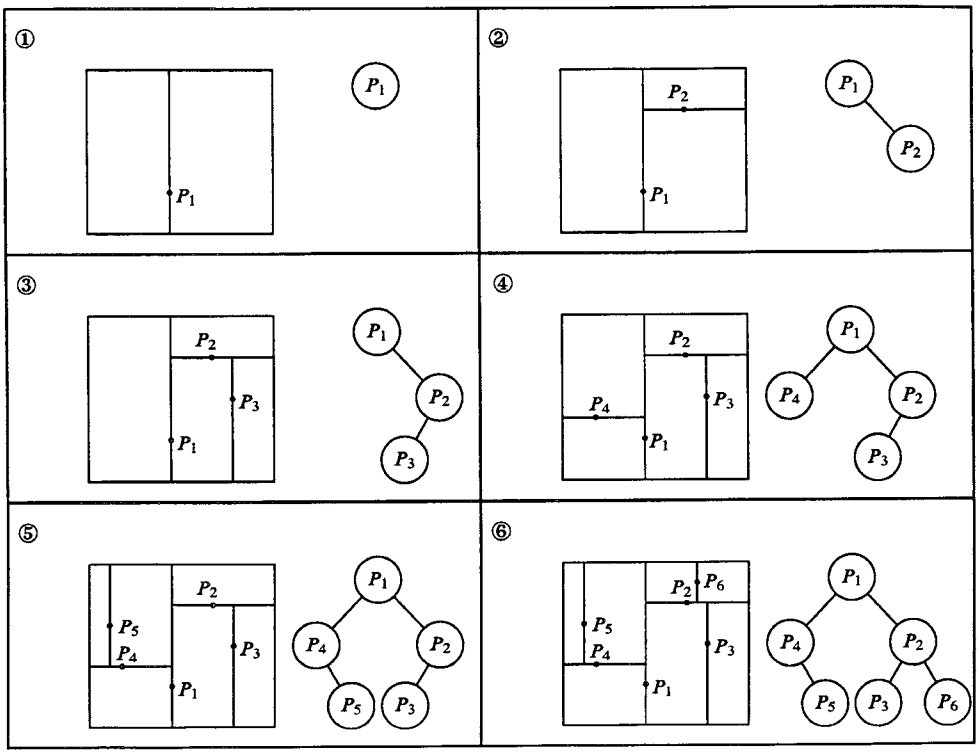

FIG. 1.1. *The stepwise constructions of a 2-d tree of six points.*

We consider the following probabilistic model for our average-case analysis. A sequence of n independently and identically distributed random points in $[0,1]^k$ is given, where each coordinate is uniformly distributed over the unit interval and is independent of other coordinates. Then we construct the k-d tree $\mathcal{T}$, called a *random k-d tree*. For partial match queries, we then consider a query of the form $\mathcal{Y} = (Y_1, \ldots, Y_k)$, where Y_j is either the unit interval (meaning don't-care) or a random variable uniformly distributed over the unit interval ($Y_j = \text{Uniform}(0,1)$), the number of specified coordinates s satisfying $0 \leq s \leq k$. We then perform the range search of

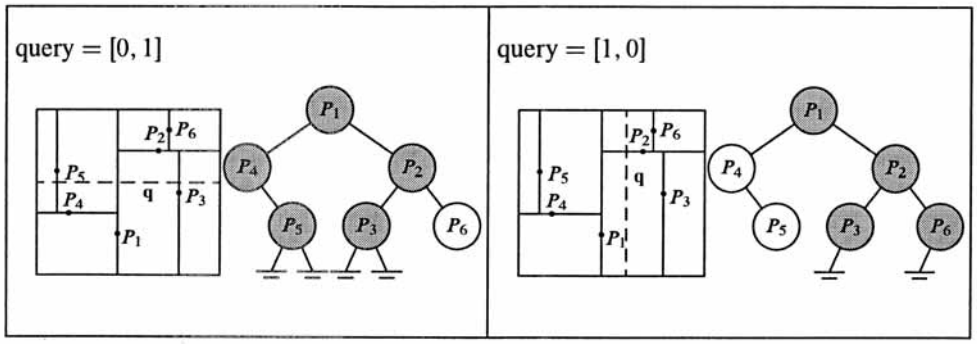

FIG. 1.2. *Two instances of partial match queries (as marked by the dashed line) in a 2-d tree: visited nodes are the filled-in grey circles. Here the query pattern* $[0,1]$ *means that the first coordinate is unspecified and the second is specified; the meaning of* $[1,0]$ *is similar.*

the query $\mathcal{Y}$ in the tree $\mathcal{T}$. The number of nodes visited is a random variable, say, X_n. The main quantity of interest in this paper is the expected value $Q_n := \mathbb{E}(X_n)$, or simply the *expected cost of a random partial match query in random k-d trees of n nodes.*

While the uniform model we are considering may seem too idealized, it is simple yet mathematically tractable; also the asymptotics obtained under such a model usually subsists under more general ones; see [17] for more discussions.

For convenience, throughout this paper we write the query pattern as $\mathbf{q} := [q_1, \dots, q_k]$ since Q_n depends only on $\mathbf{q}$, where $q_j \in \{0,1\}$ and $0 \le q_1 + \cdots + q_k \le k$. Here $q_j = 0$ means that the jth coordinate is unspecified, and $q_j = 1$ means that it is specified. Flajolet and Puech [17] showed that

$$Q_n \sim C n^{\alpha-1} \qquad (1 \le s < k), \tag{1.1}$$

where C is a constant and $\alpha > 1$ solves the equation

$$(\alpha + q_1) \cdots (\alpha + q_k) = 2^k,$$

or, equivalently, the equation $\alpha^{k-s}(\alpha+1)^s = 2^k$, where $s = q_1 + \cdots + q_k$ denotes the number of specified coordinates. Their result corrected the original claim of Bentley [2, p. 513] that $Q_n = O(n^{1-s/k})$ since α can be written as $\alpha = 1 - s/k + \varepsilon(s/k)$, where $\varepsilon(u) > 0$ for $0 < u < 1$; see [17] and [6].

The approach of Flajolet and Puech [17] is based on a linear differential system and starts from the generating function $y_1(z) := \sum_{n\ge1}(n+1)Q_n z^n$. Then

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathbf{y}(z) = \Omega(z)\mathbf{y}(z) + \mathbf{b}(z),$$

where $\mathbf{y}(z) = (y_1(z), \dots, y_{2k-s}(z))^T$, $\mathbf{b}(z) = (b_1(z), \dots, b_{2k-s}(z))^T$, and $\Omega(z)$ is a $(2k-s) \times (2k-s)$ matrix. For example, when $\mathbf{q} = [0,1,1]$,

$$\frac{\mathrm{d}}{\mathrm{d}z}\begin{pmatrix} y_1(z) \\ y_2(z) \\ y_3(z) \\ y_4(z) \end{pmatrix} = \begin{pmatrix} \frac{1}{z(1-z)} & \frac{2}{1-z} & 0 & -\frac{1}{z^2(1-z)} \\ 0 & 0 & \frac{2}{1-z} & 0 \\ \frac{2}{1-z} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} y_1(z) \\ y_2(z) \\ y_3(z) \\ y_4(z) \end{pmatrix} + \begin{pmatrix} \frac{1}{(1-z)^3} \\ \frac{2}{(1-z)^3} \\ \frac{2}{(1-z)^3} \\ 0 \end{pmatrix}.$$

From this system, they proved that the generating function $Q(z) := \sum_{n\geq 1} Q_n z^n$ satisfies

$$Q(z) \sim K(1-z)^{-\alpha}$$

for some constant K and for z in some region in the z-plane. Then (1.1) follows from singularity analysis (see [16]). But as is often the case for linear systems, the constant K or C in (1.1) does not seem to have simple computable forms by this approach.

Chanzy, Devroye, and Zamora-Cura [6] proved (1.1) by a more combinatorial and probabilistic approach, but the calculation of C still remains open.

The recurrence for Q_n we are solving is of the form ($j_0 := n$)

$$\begin{aligned} Q_n = 1 &+ \sum_{1\leq i<k} 2^i \sum_{1\leq j_i<\cdots<j_1<n} \prod_{1\leq r\leq i} \frac{1+q_r j_r}{j_{r-1}(1+q_r j_{r-1})} \\ &+ 2^k \sum_{1\leq j_k<\cdots<j_1<n} Q_{j_k} \prod_{1\leq r\leq k} \frac{1+q_r j_r}{j_{r-1}(1+q_r j_{r-1})} \end{aligned} \tag{1.2}$$

for $n \geq 1$, with $Q_0 := 0$. In particular (writing $Q_n = Q_n[q_1, \ldots, q_k]$)

$$\begin{aligned} Q_n[0,1] &= 3 - \frac{2}{n} + \frac{4}{n^2} \sum_{1\leq j<n} (n-1-j) Q_j[0,1], \\ Q_n[0,0,1] &= 7 - \frac{6}{n} - \frac{4}{n} H_{n-1} \\ &\quad + \frac{8}{n} \sum_{1\leq j<n} \left(H_{n-1} - H_j - (j+1)\left(H^{(2)}_{n-1} - H^{(2)}_j \right) \right) Q_j[0,0,1] \end{aligned}$$

for $n \geq 1$ with $Q_0 := 0$. Here $H^{(i)}_n := \sum_{1\leq j\leq n} j^{-i}$ and $H_n = H^{(1)}_n$.

While the problem we are studying in this paper is similar to partial match queries in random quadtrees (see [14, 9]), the nature of the associated analytic problems is very different (although both problems can be written in terms of linear systems). Intuitively, random quadtrees seem to have more independence in subtrees, while the independence of k-d trees is somehow "bound" by the cyclic construction. For partial match queries in quadtrees, the dominant asymptotic approximation to the expected cost depends only on the number (but not the pattern) of specified coordinates; in contrast, the leading constant in the asymptotic approximation to the expected cost for k-d trees depends strongly on the query pattern, making the analytic problem harder.

We propose in the next section a more straightforward approach, which is based on translating the system of recurrences satisfied by Q_n into a *scalar differential equation with polynomial coefficients* for the generating function of Q_n. Some explicitly solvable cases of the differential equation are discussed in section 3. We then solve the general differential equation by using Mellin transforms (see Flajolet, Gourdon, and Dumas [15]) in section 4. This Mellin approach leads, by suitably changing the initial conditions, to a series form for K that is absolutely convergent. But the justification of the application of the Mellin inversion integrals is more delicate and requires stronger analytic estimates. We thus develop in section 5 a different approach based on extending our asymptotic theory for Cauchy–Euler differential equations (see [10]). Such an approach is very general and is mostly algebraic and elementary in nature, requiring little knowledge in differential equations as in [10]. It is also applicable to

other situations, where one has linear differential equations with polynomial coefficients. We briefly indicate the application to k-d-t trees in section 6, which are locally balanced versions of k-d trees; see [11]. The development of a general theory with many applications will be given elsewhere; see [7] for further exploration of the same approach to quadtrees.

Notation. Throughout this paper, $[q_1, \dots, q_k] \in \{0,1\}^k$, $k \geq 2$, always denotes the query pattern with $s := q_1 + \cdots + q_k$, where 0 means unspecified and 1 specified. We also introduce the symbol

$$\rho_{n,j}(q) := \frac{1+qj}{1+qn} \qquad (q \in \{0,1\}).$$

Finally, $\alpha > 1$ denotes the zero of the equation (except in section 6 on k-d-t trees)

$$(x+q_1)\cdots(x+q_k) - 2^k = x^{k-s}(x+1)^s - 2^k = 0.$$

2. From recurrences to differential equations. In this section, we first show that the expected cost $Q_n = Q_n[q_1, \dots, q_k]$ satisfies a system of recurrences (of quicksort type) and then derive a "normal form" for the differential equation satisfied by the generating function $Q(z) = Q[q_1, \dots, q_k](z) := \sum_{n\geq 1} Q_n z^n$.

Basic recurrences. By definition, $Q_0 = 0$ for all query patterns.

LEMMA 2.1. *The expected cost Q_n satisfies the recurrence*

$$Q_n[q_1, \dots, q_k] = 1 + \frac{2}{n} \sum_{1\leq j<n} \rho_{n,j}(q_1) Q_j[q_2, \dots, q_k, q_1] \qquad (n \geq 1), \tag{2.1}$$

with $Q_0 = 0$.

Proof. When the first coordinate is unspecified, the search is conducted in both subtrees. The probability that the left subtree is of size j is $1/n$ for $0 \leq j < n$. By cyclic construction of k-d trees, we then have

$$Q_n[q_1, \dots, q_k] = 1 + \frac{1}{n} \sum_{0\leq j<n} \left(Q_j[q_2, \dots, q_k, q_1] + Q_{n-1-j}[q_2, \dots, q_k, q_1]\right);$$

thus (2.1) follows with $q_1 = 0$. On the other hand, if $q_1 = 1$, then the probability of going to the left subtree of the root (when it has j nodes) is $(j+1)/(n+1)$ since with probability 1 the query results in an unsuccessful search. Thus

$$Q_n[q_1, \dots, q_k] = 1 + \frac{1}{n} \sum_{0\leq j<n} \left(\frac{j+1}{n+1} Q_j[q_2, \dots, q_k, q_1] + \frac{n-j}{n+1} Q_{n-1-j}[q_2, \dots, q_k, q_1]\right).$$

This proves (2.1). ∎

Write $Q_n^{[j]} := Q_n[q_j, \dots, q_k, q_1, \dots, q_{j-1}]$, $1 \leq j \leq k$.

COROLLARY 2.2. *The $Q_n^{[j]}$'s satisfy the system of recurrences*

$$\begin{cases} Q_n^{[1]} = 1 + \dfrac{2}{n} \displaystyle\sum_{1\leq j<n} \rho_{n,j}(q_1) Q_j^{[2]}, \\ \qquad\vdots \qquad\qquad\qquad (n \geq 1). \\ Q_n^{[k]} = 1 + \dfrac{2}{n} \displaystyle\sum_{1\leq j<n} \rho_{n,j}(q_k) Q_j^{[1]}, \end{cases} \tag{2.2}$$

Proof. The proof follows from Lemma 2.1 and the cyclic relation of k-d trees. □

The recurrence (1.2) then follows from (2.2).

COROLLARY 2.3. *If we assume that* $Q_n^{[1]} \sim Cn^{\beta-1}$ *for* $\beta > 1$, *then*

$$Q_n^{[j]} \sim \frac{2^{k-j+1}C}{(\beta+q_j)\cdots(\beta+q_k)}\, n^{\beta-1} \qquad (j=1,\dots,k), \tag{2.3}$$

and β *solves the equation* $(\beta+q_1)\cdots(\beta+q_k)=2^k$.

Proof. From (2.2), we have

$$Q_n^{[k]} \sim \frac{2C}{n^{1+q_k}} \sum_{1\le j<n} j^{\beta-1+q_k} \sim \frac{2C}{\beta+q_k}\, n^{\beta-1}.$$

Then the asymptotics of $Q_n^{[k-j]}$ can be obtained one after another. On the other hand, since

$$Q_n^{[1]} \sim \frac{2^k C}{(\beta+q_1)\cdots(\beta+q_k)}\, n^{\beta-1} \sim Cn^{\beta-1},$$

β has to satisfy the equation $(\beta+q_1)\cdots(\beta+q_k)=2^k$. Thus $\beta=\alpha$. □

The corollary says that once we obtain the asymptotics of one of the $Q_n^{[j]}$'s, those of the others are all known. One can show that the number of distinct query patterns needed to be computed in order to derive Q_n for all 2^k possible queries is equal to (essentially the number of different types of necklaces with two distinct colors)

$$N_k := k^{-1} \sum_{d|k} \phi(d) 2^{k/d}, \tag{2.4}$$

where $\phi(d)$ denotes Euler's totient function; see Flajolet and Sedgewick [18, p. 18]. For large k, $N_k \sim 2^k/k$.

We can rewrite and simplify (2.2) as follows.

COROLLARY 2.4. *For* $n \ge 1$,

$$\begin{pmatrix} Q_n^{[1]} \\ \vdots \\ Q_n^{[k]} \end{pmatrix} = \mathbf{M} \begin{pmatrix} Q_{n-1}^{[1]} \\ \vdots \\ Q_{n-1}^{[k]} \end{pmatrix} + \begin{pmatrix} \frac{1+q_1}{n+q_1} \\ \vdots \\ \frac{1+q_k}{n+q_k} \end{pmatrix}, \tag{2.5}$$

where $\mathbf{M} = (a_{i,j})_{k\times k}$ *with* $a_{i,i} = (n-1)/(n+q_i)$, $a_{i,(i+1)\bmod k} = 2/(n+q_i)$, *and all other entries are zeros.*

Proof. Taking the difference $n(n+q_j)Q_n^{[j]} - (n-1)(n-1-q_j)Q_{n-1}^{[j]}$ gives

$$(n+q_j)Q_n^{[j]} = (n-1)Q_{n-1}^{[j]} + 2Q_{n-1}^{[j+1]} + 1 + q_j \qquad (1\le j\le k),$$

where $Q_n^{[k+1]} = Q_n^{[1]}$. □

The nonhomogeneous matrix recurrence (2.5) can be converted into a homogeneous one by considering

$$\bar{Q}_n^{[j]} := Q_n^{[j]} + \frac{1+\sum_{1\le\ell<k}(2-q_1)\cdots(2-q_\ell)}{2^s-1};$$

then

$$\begin{pmatrix} \bar{Q}_n^{[1]} \\ \vdots \\ \bar{Q}_n^{[k]} \end{pmatrix} = \mathbf{M} \begin{pmatrix} \bar{Q}_{n-1}^{[1]} \\ \vdots \\ \bar{Q}_{n-1}^{[k]} \end{pmatrix},$$

with obvious initial conditions.

Integral equations. Lemma 2.1 is next translated into an integral equation for the generating function $Q(z)$. For convenience, we introduce two integral operators,

$$\begin{cases} I_0[\varphi](z) := \displaystyle\int_0^z \frac{\varphi(t)}{1-t}\,\mathrm{d}t, \\ I_1[\varphi](z) := \displaystyle\frac{1-z}{z}\int_0^z \frac{t\varphi(t)}{(1-t)^2}\,\mathrm{d}t. \end{cases}$$

Let $f_j(z) = Q[q_j, \dots, q_k, q_1, \dots, q_{j-1}](z)$, $1 \le j \le k$.

LEMMA 2.5. *The generating function f_1 satisfies the integral equation*

$$f_1(z) = \frac{z}{1-z} + 2I_{q_1}[f_2](z). \tag{2.6}$$

Proof. The proof for (2.6) with $q_1 = 0$ is straightforward. When $q_1 = 1$, (2.6) follows from

$$\begin{aligned} \sum_{n\ge1} \frac{z^n}{n(n+1)} \sum_{0\le j<n} (j+1)Q_j &= \sum_{n\ge1} \left(\frac{z^n}{n} - \frac{z^n}{n+1}\right) \sum_{0\le j<n} (j+1)Q_j \\ &= \int_0^z \frac{1}{1-t}\,(tQ(t))'\,\mathrm{d}t - \frac{1}{z}\int_0^z \frac{t}{1-t}\,(tQ(t))'\,\mathrm{d}t. \end{aligned}$$

□

The form given in (2.6) for $q_1 = 1$ is the main diverging point on which our approach differs from that used in [17]. Roughly, we keep the inverse operator of I_1 to be of first order, so that the associated linear system is of degree k, instead of $2k - s$ as used in [17]; cf. also (2.5).

COROLLARY 2.6. *We have*

$$\begin{cases} f_1(z) = 2I_{q_1}[f_2](z) + \dfrac{z}{1-z}, \\ \qquad\vdots \\ f_k(z) = 2I_{q_k}[f_1](z) + \dfrac{z}{1-z}. \end{cases}$$

By composing the above integral equations, we obtain a single integral equation for $f_1(z) = Q(z)$.

COROLLARY 2.7. *The generating function of Q_n satisfies*

$$Q(z) - 2^k(I_{q_1} \circ \cdots \circ I_{q_k})[Q](z) = G_0(z), \tag{2.7}$$

where $(I_{q_1} \circ \cdots \circ I_{q_j})[f](z) := I_{q_1}[\cdots[I_{q_j}[f]]\cdots](z)$, $1 \le j \le k$, *and*

$$G_0(z) = \frac{z}{1-z} + \sum_{1\le j<k} 2^j (I_{q_1} \circ \cdots \circ I_{q_j}) \left[\frac{z}{1-z}\right](z).$$

Note that $Q_n = [z^n]G_0(z)$ for $1 \le n < k$, where $[z^n]\varphi(z)$ denotes the coefficient of z^n in the Taylor expansion of φ. On the other hand, by (1.2),

$$Q_n = 1 + \sum_{1\le i\le n} 2^i \sum_{1\le j_i<\cdots<j_1<n} \prod_{1\le r\le i} \frac{1+q_r j_r}{j_{r-1}(1+q_r j_{r-1})} \qquad (1 \le n < k).$$

Differential equations. Define the differential operator $\vartheta := (1-z)(\mathrm{d}/\mathrm{d}z)$. Then the inverse operator of I_q is

$$I_q^{-1} = \vartheta + \frac{q}{z} \qquad (q \in \{0,1\}),$$

and the integral equation (2.7) can be converted into a scalar differential equation

(2.8)
$$\left(\vartheta + \frac{q_k}{z}\right)\cdots\left(\vartheta + \frac{q_1}{z}\right) Q(z) - 2^k Q(z) = \sum_{1\le j\le k} 2^{j-1}\left(\vartheta + \frac{q_k}{z}\right)\cdots\left(\vartheta + \frac{q_j}{z}\right)\frac{z}{1-z}.$$

If we multiply both sides by z^k, then the left-hand side can be written in the form $\sum_{0\le j\le k}(1-z)^j L_{k,j}(\vartheta)Q(z)$ for some polynomials $L_{k,j}$, all of degree k. But this approach does not lead to simpler solutions.

Two special cases are indicative of the general pattern for simplification.

(i) If all q_j's are zeros, then (2.8) becomes

$$\left(\vartheta^k - 2^k\right) Q(z) = \frac{2^k - 1}{1-z}. \tag{2.9}$$

(ii) If the q_j's are all ones, then, by the relation

$$\left(\vartheta + \frac{1}{z}\right)\frac{\varphi(z)}{z^j} = \frac{1-j(1-z)}{z^{j+1}}\,\varphi(z) \qquad (j = 0, 1, \dots),$$

we have

$$\left(\vartheta - 2^k + \frac{(\vartheta+1)^k - \vartheta^k}{z}\right) Q(z) = z^{-1}\left(2 - 2^{k+1} + \frac{k2^k}{1-z}\right),$$

which, by multiplying both sides by z, yields

$$\left((\vartheta+1)^k - 2^k\right) Q(z) - (1-z)\left(\vartheta^k - 2^k\right) Q(z) = 2 - 2^{k+1} + \frac{k2^k}{1-z}. \tag{2.10}$$

These observations suggest that we define a "minimum exponent" $\mu_{\mathbf{q}}$ as $k-v+1$, where v denotes the last position of the first block of 1's (from left to right) and $\mu_{\mathbf{q}} := 0$ if $s = 0$. In symbols,

$$[q_1, \dots, q_k] = [0, \dots, 0, 1, \dots, \underbrace{1, 0, *, \dots, *}_{\mu_{\mathbf{q}}}] \qquad (* \in \{0,1\})$$

so that $\mu_{\mathbf{q}} := 1$ if $s = k$.

LEMMA 2.8. *Let $\mu_{\mathbf{q}}$ be defined as above. Then*

$$\begin{aligned} z^{\mu_{\mathbf{q}}}&\left(\left(\vartheta + \frac{q_k}{z}\right)\cdots\left(\vartheta + \frac{q_1}{z}\right) - 2^k\right) Q(z) \\ &= P_{k,0}(\vartheta)Q(z) - \sum_{1\le \ell\le \mu_{\mathbf{q}}} (1-z)^\ell P_{k,\ell}(\vartheta)Q(z), \end{aligned} \tag{2.11}$$

where the $P_{k,\ell}$'s are polynomials of degree k defined by $P_{k,0}(x) = p_{k,0}(x) - 2^k$, and for $1 \le \ell \le \mu_{\mathbf{q}}$,

$$P_{k,\ell}(x) := p_{k,\ell}(x) + (-1)^\ell 2^k \binom{\mu_{\mathbf{q}}}{\ell}, \tag{2.12}$$

with

$$\begin{cases} p_{k,1}(x) = (x+q_k-1)p_{k-1,1}(x) + (x+\mu_{\mathbf{q}}-1)p_{k-1,0}(x), \\ p_{k,\ell}(x) = (x+q_k-\ell)p_{k-1,\ell}(x) - (x+\mu_{\mathbf{q}}-\ell)p_{k-1,\ell-1}(x) \qquad (2 \le \ell < \mu_{\mathbf{q}}), \end{cases}$$

and the boundary values $p_{k,0}(x) = x^{k-s}(x+1)^s$ and $p_{k,\mu_{\mathbf{q}}}(x) = (-1)^{\mu_{\mathbf{q}}+1}x^k$ for $k \ge 1$.

Proof. The two cases of when $s = 0$ and when $s = k$ are easily checked by (2.9) and (2.10), respectively. We assume that $1 \le s < k$.

For $k = 2$, there are two query patterns, $[0,1]$ and $[1,0]$. The differential equation for $[0,1]$ is ($\mu_{\mathbf{q}} = 1$)

$$\left(\vartheta(\vartheta+1) - 4 - (1-z)(\vartheta^2-4)\right) Q(z) = 1 + \frac{6z}{1-z}, \tag{2.13}$$

and that for $[1,0]$ is ($\mu_{\mathbf{q}} = 2$)

$$\left((\vartheta(\vartheta+1)-4) - (1-z)\left(2\vartheta^2+\vartheta-7\right) + (1-z)^2(\vartheta^2-4)\right) Q(z) = \frac{4z^2}{1-z}. \tag{2.14}$$

Thus (2.11) holds.

When the query is of the form

$$[\underbrace{0,\dots,0}_{k-s},\underbrace{1,\dots,1}_{s}] \qquad (1 \le s < k), \tag{2.15}$$

we have $\mu_{\mathbf{q}} = 1$, and the left-hand side of (2.8) satisfies

$$z\left(\vartheta + \frac{1}{z}\right)^s \vartheta^{k-s} Q(z) = \left(\vartheta^{k-s}(\vartheta+1)^s - 2^k - (1-z)\left(\vartheta^k - 2^k\right)\right) Q(z);$$

thus (2.11) holds.

The remaining cases follow by induction on k. The proof is messy and omitted here. □

Similarly, the right-hand side of (2.8) can be rewritten as

$$z^{\mu_{\mathbf{q}}} \sum_{1 \le j \le k} 2^{j-1} \left(\vartheta + \frac{q_k}{z}\right) \cdots \left(\vartheta + \frac{q_j}{z}\right) \frac{z}{1-z} = \sum_{-1 \le \ell < \mu_{\mathbf{q}}} \nu_{k,\ell}(1-z)^\ell. \tag{2.16}$$

For the special query pattern (2.15), we have

$$\begin{cases} \nu_{k,-1} = (s+1)2^k - 2^s, \\ \nu_{k,0} = -2^{k+1} + 2^{k-s} + 1. \end{cases} \tag{2.17}$$

For other cases, we obtain, by induction,

$$\nu_{k,\ell} := \begin{cases} \left(2^{k-1} + \nu_{k-1,-1}\right)(q_k+1), & \ell = -1, \\ 2^{k-1}(q_k+1)(-1)^\ell \binom{\mu_{\mathbf{q}}}{\ell+1} & \\ \qquad + \left((\ell-\mu_{\mathbf{q}})\nu_{k-1,\ell-1} + (q_k-\ell)\nu_{k-1,\ell}\right), & 0 \le \ell \le \mu_{\mathbf{q}}-2, \\ 2^{k-1}(q_k+1)(-1)^{\mu_{\mathbf{q}}} - \nu_{k-1,\mu_{\mathbf{q}}-2}, & \ell = \mu_{\mathbf{q}}-1, \end{cases} \tag{2.18}$$

with the initial values $\nu_{\tau+1,-1} = 2(2^{\tau+1} - 1)$ and $\nu_{\tau+1,0} = -3 \cdot 2^{\tau} + 1$, τ being the largest integer for which $q_1 = \cdots = q_\tau = 0$ but $q_{\tau+1} = 1$. The proof is again lengthy and laborious, and thus omitted here.

Combining (2.11) and (2.16), we obtain the following "normal form" for Q.

PROPOSITION 2.9. *The generating function $Q(z)$ satisfies the initial value problem*

$$\begin{cases} P_{k,0}(\vartheta)Q(z) - \displaystyle\sum_{1\le\ell\le\mu_{\mathbf{q}}} (1-z)^{\ell} P_{k,\ell}(\vartheta)Q(z) = \sum_{-1\le\ell<\mu_{\mathbf{q}}} \nu_{k,\ell}(1-z)^{\ell}, \\ Q(z) = z + Q_2 z^2 + \cdots + Q_{k-1}z^{k-1} + \cdots, \end{cases} \tag{2.19}$$

where the $P_{k,\ell}$'s are polynomials of degree k defined in (2.12) *and the $\nu_{k,\ell}$'s are constants given in* (2.17) *and* (2.18).

The exact solution for such a general problem is still not obvious and exists only for a few special cases to be discussed below.

3. Explicitly solvable cases. Exactly solvable cases of (2.19) include $s = 0$, $s = k$, and $k = 2$, where s is the number of specified coordinates.

None specified: $s = 0$. When no coordinate is specified, the search cost is obviously n. The associated differential equation is given by (2.9), with the initial conditions $Q(z) = z + 2z^2 + \cdots + (k-1)z^{k-1} + \cdots$. The solution is $Q(z) = z/(1-z)^2$ for all $k \ge 1$. This is a case when *the particular solution is itself the exact solution because all initial values coincide.*

All queries specified: $s = k$. In this case, the two Q_n's on the two sides of the recurrence (2.1) are the same so that Q_n satisfies

$$Q_n = 1 + \frac{2}{n(n+1)} \sum_{1\le j<n} (j+1)Q_j \qquad (n \ge 1),$$

with $Q_0 = 0$, which is easily solved to be

$$Q_n = 2(H_{n+1} - 1) \qquad (n \ge 1). \tag{3.1}$$

The solution is invariant in k. This is nothing but the expected number of comparisons used for an unsuccessful search in random binary search trees; see [6, 23].

Note that $Q(z)$ satisfies the differential equation (2.10), with the initial values $Q(0) = 0$ and

$$Q_j = 1 + [z^{j-2}]\frac{1}{(1-z)^2}\left(1 - \frac{2\Gamma(j+2z)}{\Gamma(j+2)\Gamma(1+2z)}\right) \qquad (1 \le j < k),$$

Γ being the Gamma function. The differential equation (2.10) can be rewritten as

$$\left((\vartheta+1)^k - 2^k\right)(zQ(z)) = 2 + (k-2)2^k + \frac{k2^k}{1-z}z,$$

which is indeed of Cauchy–Euler type (see [10]). The exact solution is

$$Q(z) = \frac{2}{z\,(1-z)} \log\left(\frac{1}{1-z}\right) - \frac{2}{1-z},$$

from which one derives (3.1). This is another case *when the particular solution is itself the exact solution.*

2-*d trees.* In this case, $Q[0,1](z)$ satisfies (2.13) and $Q[1,0](z)$ satisfies (2.14), both with the initial conditions $Q(0)=0$ and $Q'(0)=1$.

Let ${}_2F_1$ denote Gauss's hypergeometric function

$$ {}_2F_1\left(\begin{matrix}a,b\\ c\end{matrix}\,\middle|\, z\right)=1+\sum_{j\ge 1}\frac{a^{\bar j}\,b^{\bar j}}{c^{\bar j}\,j!}\,z^j, $$

where $x^{\bar j}=x(x+1)\cdots(x+j-1)$ denotes the rising factorial.

LEMMA 3.1 (exact solutions for $Q(z)$). *The generating function of $Q_n[0,1]$ satisfies*

$$ Q[0,1](z)=\frac{13}{2}(1-z)^{-\alpha}{}_2F_1\left(\begin{matrix}2-\alpha,-2-\alpha\\ 1\end{matrix}\,\middle|\, z\right)-\frac{7}{2}+4z-\frac{3}{2}z^2-\frac{3}{1-z}, \tag{3.2} $$

and that for $Q_n[1,0]$

$$ Q[1,0](z)=\frac{13}{2}z(1-z)^{-\alpha}{}_2F_1\left(\begin{matrix}3-\alpha,-1-\alpha\\ 3\end{matrix}\,\middle|\, z\right)+2-\frac{7}{2}z+\frac{3}{2}z^2-\frac{2}{1-z}, \tag{3.3} $$

where $\alpha=(\sqrt{17}-1)/2$.

The corresponding recurrence relations have the forms

$$ \begin{aligned} Q_n[0,1]&=3-\frac{2}{n}+\frac{4}{n^2}\sum_{1\le j<n}(n-1-j)Q_j[0,1],\\ Q_n[1,0]&=3-\frac{2}{n}+\frac{4}{n(n+1)}\sum_{1\le j<n}(n-1-j+H_{n-1}-H_j)Q_j[1,0], \end{aligned} $$

with $Q_0=0$ for both cases.

COROLLARY 3.2 (exact solutions for Q_n). *For $n\ge 3$,*

$$ \begin{aligned} Q_n[0,1]&=\frac{13}{2}\sum_{0\le j\le n}\binom{n-j+\alpha-1}{n-j}\frac{(2-\alpha)^{\bar j}(-2-\alpha)^{\bar j}}{j!j!}-3,\\ Q_n[1,0]&=13\sum_{0\le j<n}\binom{n-j+\alpha-2}{n-1-j}\frac{(3-\alpha)^{\bar j}(-1-\alpha)^{\bar j}}{(j+2)!j!}-2. \end{aligned} $$

Proof. The proof follows from taking coefficients of z^n on both sides of (3.2) and of (3.3). □

COROLLARY 3.3 (asymptotics of Q_n).

$$ \begin{aligned} Q_n[0,1]&=\frac{13(2\alpha-3)}{2}\cdot\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}\,n^{\alpha-1}-3+O(n^{\alpha-2}),\\ Q_n[1,0]&=\frac{13(8-5\alpha)}{2}\cdot\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}\,n^{\alpha-1}-2+O(n^{\alpha-2}). \end{aligned} $$

Proof. The proof follows from applying the exact solutions of Q_n or by applying singularity analysis to (3.2) and (3.3), and then using Gauss's identity (see [1, Eq. 15.1.20, p. 556]). □

Numerically,

$$ \frac{13(2\alpha-3)}{2}\cdot\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}\approx 2.55275,\qquad \frac{13(8-5\alpha)}{2}\cdot\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}\approx 1.99312, $$

which are to be compared with the asymptotic approximation

$$\frac{1}{2}\cdot\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^3}\,n^{\alpha-1}\approx 1.59099\,n^{\alpha-1}$$

for the expected cost of partial match queries in random two-dimensional quadtrees; see [9, 14].

Proof of Lemma 3.1. Consider first the case when the query pattern is $[0,1]$. In this case, the differential equation is

$$z(1-z)^2Q''(z)+(1-z)^2Q'(z)-4\,zQ(z)=-5+\frac{6}{1-z}.$$

To solve this equation, we first observe that the particular solution is given by

$$y_p(z)=-\frac{7}{2}+4z-\frac{3}{2}z^2-\frac{3}{1-z}.$$

Thus we consider $Q(z)=(1-z)^{-\alpha}y(z)+y_p(z)$, and we have

$$z(1-z)^2y''(z)+(1-z)(1-z+2\,z\alpha)y'(z)+\left(\alpha+z\alpha^2-4\,z\right)y(z)=0.$$

Since $\alpha(\alpha+1)-4=0$, we can simplify the above equation and obtain

$$z(1-z)y''(z)+(1+(2\alpha-1)z)y'(z)+\alpha y(z)=0.$$

This equation is then rewritten as a generalized hypergeometric differential equation

$$\left(\delta^2-z(\delta^2-2\alpha\delta-\alpha)\right)y(z)=0,$$

where $\delta=z(\mathrm{d}/\mathrm{d}z)$. The differential equation has two fundamental solutions, and the only one that is regular at the origin is

$${}_2F_1\left(\begin{array}{c}2-\alpha,-2-\alpha\\1\end{array}\,\middle|\,z\right).$$

Thus the solution is of the form

$$Q(z)=-\frac{7}{2}+4z-\frac{3}{2}z^2-3(1-z)^{-1}+c_0(1-z)^{-\alpha}{}_2F_1\left(\begin{array}{c}2-\alpha,-2-\alpha\\1\end{array}\,\middle|\,z\right),$$

and by matching the initial values, we conclude (3.2).

For the query pattern $[1,0]$, the particular solution for (2.14) is

$$y_p(z):=2-\frac{7}{2}z+\frac{3}{2}z^2-\frac{2}{1-z}.$$

The only different part of the above proof is considering $y(z):=\left(Q[1,0](z)-y_p(z)\right)/z$; the remaining analysis is similar and omitted. □

The approach we used for deriving (3.2) and (3.2) is similar to the one used in [14]. However, such an approach does not apply for $k\geq 3$. We need a different approach to solving (2.19).

4. Constants for general query patterns. We rewrite the differential equation (2.19) as

$$P_0(\vartheta)Q(z) = \sum_{1\le \ell\le \mu} (1-z)^\ell P_\ell(\vartheta)Q(z) + \sum_{-1\le \ell<\mu} \nu_{k,\ell}(1-z)^\ell, \tag{4.1}$$

where, for simplicity, we write $P_\ell = P_{k,\ell}$ (defined in (2.12)) and $\mu = \mu_{\mathbf{q}}$.

Our main result in this paper is the following asymptotic estimate for Q_n.

THEOREM 4.1. *The expected cost of a random partial match query in a random k-d tree of n nodes satisfies*

$$Q_n \sim \frac{K}{\Gamma(\alpha)}\, n^{\alpha-1} \qquad (1 \le s < k), \tag{4.2}$$

where the constant K is defined in (4.6) *below.*

This theorem is proved in the next section by extending the asymptotic theory in [10] for Cauchy–Euler differential equations. The idea is roughly as follows. If the right-hand side of (4.1) is independent of Q, then the solution of Q would satisfy $Q(z) \sim c_1(1-z)^{-\alpha} + y_p(z)$ for some constant c_1 and particular solution $y_p(z)$; see [10]. But the intricate part here is that the right-hand side of (4.1), expected to have a smaller contribution to $Q(z)$, itself depends on Q.

In this section we give an approach to the proof of (4.2) based on Mellin-type integrals. The proof is almost complete, up to an estimate in the complex plane that is needed in justifying the Mellin inversion integrals. The advantage of this approach is that it quickly gives the right form for the constant.

Shifting the initial values. A minor but crucial step in our analysis is to consider the function $f(z) := Q(z) - \sum_{1\le j<k} Q_j z^j$. The differential equation for f remains the same, but all initial values become zero. Indeed, by the integral equation (2.7), we have

$$f(z) - 2^k (I_{q_1} \circ \cdots \circ I_{q_k})[f](z) = G_1(z),$$

with $f(0) = \cdots = f^{(k-1)}(0) = 0$, where

$$G_1(z) := G_0(z) - \sum_{1\le j<k} Q_j z^j - 2^k (I_{q_1} \circ \cdots \circ I_{q_k}) \left[\sum_{1\le j<k} Q_j z^j \right](z).$$

Note that $I_q[z^j] = z^{j+1}/(j+1+q) + \cdots$, $q \in \{0,1\}$. This, together with the relations $Q_j = [z^j]G_0(z)$ for $0 \le j < k$, implies that $[z^j]G_1(z) = 0$ for $j < k$.

Now applying the inverse operators to the above integral equation, and then multiplying both sides by z^μ, we obtain

$$P_0(\vartheta)f(z) = \sum_{1\le \ell\le \mu} (1-z)^\ell P_\ell(\vartheta)f(z) + g(z), \tag{4.3}$$

with $f^{(j)}(0) = 0$, $0 \le j \le k-1$, where

$$\begin{aligned} g(z) = {} & z^\mu \sum_{1\le j\le k} 2^{j-1} \left(\vartheta + \frac{q_k}{z}\right) \cdots \left(\vartheta + \frac{q_j}{z}\right) \frac{z}{1-z} \\ & - z^\mu \left(\left(\vartheta + \frac{q_k}{z}\right) \cdots \left(\vartheta + \frac{q_1}{z}\right) - 2^k \right) (Q_1 z + \cdots + Q_{k-1} z^{k-1}). \end{aligned}$$

By using the same arguments as in the proof of Proposition 2.9, we obtain the alternative form for g,

$$g(z) = \frac{\nu_{k,-1}}{1-z} + \Pi(z),$$

with $\Pi(z)$ being a polynomial of degree $k+\mu-1$. With this form, we can now define

$$g^*(w) := \int_0^1 (1-x)^{w-1} g(x)\,\mathrm{d}x \qquad (\Re(w) > 0). \tag{4.4}$$

Then g^* satisfies

$$g^*(w) = \frac{\nu_{k,-1}}{w-1} + \sum_{0\le j<k+\mu} \frac{(-1)^j \Pi^{(j)}(1)}{j!(w+j)};$$

see Table 5.2. The function g^* can be viewed either as the Mellin transform (see [15]) of $g(x)$ (defined to be zero for $x > 1$) or as the factorial series of the sequence $[z^j]g(z)$.

The following estimate is needed in proving the absolute convergence of the series representation (4.6) of K.

LEMMA 4.2. *The function g^* satisfies the estimate*

$$g^*(w) = O(|w|^{-\mu-1}) \tag{4.5}$$

for large $|w|$ and $|\arg(w)| \le \pi - \varepsilon$, $\varepsilon > 0$.

Proof. Observe first that by definition,

$$g(z) = z^\mu \left(\vartheta + \frac{q_k}{z}\right) \cdots \left(\vartheta + \frac{q_j}{z}\right) G_1(z),$$

and that $[z^j]G_1(z) = 0$ for $j < k$. Since each operator $\vartheta + q_k/z$ has the effect of decreasing the powers of monomials by 1 (from z^j to z^{j-1}), we can expand g as $g(z) = \sum_{j\ge\mu} g_j z^j$ for some coefficients g_j. By using this expansion and interchanging the summation and integral, we obtain

$$g^*(w) = \sum_{j\ge\mu} \frac{g_j j!}{w(w+1)\cdots(w+j)} \qquad (\Re(w) > 1),$$

which is not only a factorial series but also an asymptotic expansion for large $|w|$. ∎

Characterization of K. Now we are ready to give an explicit form for the constant K in Theorem 4.1.

PROPOSITION 4.3. *The constant K in (4.2) is given by*

$$K = \frac{1}{P_0'(\alpha)} \sum_{j\ge0} g^*(\alpha+j) B_j, \tag{4.6}$$

with the series being absolutely convergent, where

$$B_j = \sum_{1\le\ell\le\mu} \frac{P_\ell(\alpha+j)}{P_0(\alpha+j)} B_{j-\ell} \qquad (j \ge 1), \tag{4.7}$$

with the initial conditions $B_0 = 1$ and $B_j = 0$ for $j < 0$.

We defer the hard part of proving (4.6) to later sections and prove here only the absolute convergence, which is a direct consequence of the estimate (4.5) for g^* and the following lemma.

LEMMA 4.4. *Let $B(z)=\sum_{j\geq 0}B_jz^j$. Then $B(z)$ satisfies the differential equation*

$$(1-z)^\mu\left(\left(\frac{q_1-\mu z}{1-z}+\alpha+z\mathbb{D}_z\right)\cdots\left(\frac{q_k-\mu z}{1-z}+\alpha+z\mathbb{D}_z\right)-2^k\right)B(z)=0, \tag{4.8}$$

where $\mathbb{D}_z:=\mathrm{d}/\mathrm{d}z$, and the sequence B_j satisfies

$$B_j=O\left(j^{\mu-q_k-1}\right). \tag{4.9}$$

Proof. The proof is adapted from that of Lemma 3 in [7]. Define the operator

$$\Psi(q_1,\dots,q_k)[\vartheta]:=\left(\vartheta+\frac{q_k}{z}\right)\cdots\left(\vartheta+\frac{q_1}{z}\right).$$

By the Cauchy integral representation,

$$B_j=\frac{1}{2\pi i}\oint(1-z)^{-j-1}B(1-z)\mathrm{d}z\qquad(j\geq 0),$$

where, here and throughout the proof, the integration contour is a sufficiently small circle around unity, we obtain

(4.10)

$$\begin{aligned}
0&=\frac{1}{2\pi i}\oint B(1-z)(1-z)^{\alpha-1}\\
&\qquad\times\left(P_0(j+\alpha)-\sum_{1\leq\ell\leq\mu}(1-z)^\ell P_\ell(j+\alpha)\right)(1-z)^{-j-\alpha}\mathrm{d}z\\
&=\frac{1}{2\pi i}\oint B(1-z)(1-z)^{\alpha-1}\left(P_0(\vartheta)-\sum_{1\leq\ell\leq\mu}(1-z)^\ell P_\ell(\vartheta)\right)(1-z)^{-j-\alpha}\mathrm{d}z\\
&=\frac{1}{2\pi i}\oint B(1-z)(1-z)^{\alpha-1}\left(z^\mu\Psi(q_1,\dots,q_k)[\vartheta]-2^kz^\mu\right)(1-z)^{-j-\alpha}\mathrm{d}z\\
&=V_j-2^k[z^j]B(z)(1-z)^\mu,
\end{aligned}$$

where

$$\begin{aligned}
V_j&:=\frac{1}{2\pi i}\oint B(1-z)(1-z)^{\alpha-1}z^\mu\Psi(q_1,\dots,q_k)[\vartheta](1-z)^{-j-\alpha}\mathrm{d}z\\
&=\frac{1}{2\pi i}\oint B(1-z)(1-z)^{\alpha-1}z^\mu\left(\vartheta+\frac{q_k}{z}\right)\Psi(q_1,\dots,q_{k-1})[\vartheta](1-z)^{-j-\alpha}\mathrm{d}z.
\end{aligned}$$

By an integration by parts (for the term corresponding to ϑ), we have

$$V_j=\frac{1}{2\pi i}\oint B_1(1-z)(1-z)^{\alpha-1}z^\mu\Psi(q_1,\dots,q_{k-1})[\vartheta](1-z)^{-j-\alpha}\mathrm{d}z,$$

where

$$B_1(1-z):=\left(\frac{q_k-\mu(1-z)}{z}+\alpha-\vartheta\right)B(1-z).$$

Repeating the same argument $k-1$ times, we obtain

$$\begin{aligned} V_j &= \frac{1}{2\pi i}\oint (1-z)^{-j-1} z^{\mu} \\ &\qquad \times \left(\frac{q_1-\mu(1-z)}{z}+\alpha-\vartheta\right)\cdots\left(\frac{q_k-\mu(1-z)}{z}+\alpha-\vartheta\right) B(1-z)\mathrm{d}z \\ &= [z^j](1-z)^{\mu}\left(\frac{q_1-\mu z}{1-z}+\alpha+z\mathbb{D}_z\right)\cdots\left(\frac{q_k-\mu z}{1-z}+\alpha+z\mathbb{D}_z\right) B(z). \end{aligned}$$

From this and (4.10), the result (4.8) follows from multiplying both sides of (4.10) and summing over all $j \geq 0$.

According to the Frobenius method (see [24]), we seek solutions to the differential equation (4.8) of the form $B(z) = (1-z)^{\zeta}\xi(1-z)$ for some $\zeta \in \mathbb{C}$ and some function $\xi(z)$ analytic at the origin. Substituting this form into (4.8), we see that the indicial equation for (4.8) is given by

$$\prod_{1\leq j\leq k} (-\mu-\zeta+k-j+q_j) = 0.$$

In particular, the dominant zero is $\zeta = -\mu + q_k$, which is a simple zero. This implies that

$$B(z) = O\left(|1-z|^{-\mu+q_k}\right),$$

and, by singularity analysis (see [16]), proves (4.9). □

Note that, by (4.7), $B(z)$ also satisfies the differential equation

$$P_0(\alpha+z\mathbb{D}_z)B(z) = \sum_{1\leq \ell\leq \mu} (1-z)^{\ell} P_{\ell}(\alpha+\ell+z\mathbb{D}_z)B(z),$$

but this form is less manageable than (4.8) for our purposes.

While the series form (4.6) may seem recursive, it is readily modified for numerical purposes; see Table 5.1.

We next give a formal proof of the formula (4.6) by Mellin integrals, which will be justified by a more algebraic procedure.

Factorial series and Mellin integrals. Define the factorial series

$$f^*(w) := \int_0^1 (1-z)^{w-1} f(z)\,\mathrm{d}z = \sum_{j\geq k} \frac{Q_j j!}{w(w+1)\cdots(w+j)}.$$

Then $f^*(w)$ is well defined in the half-plane $\Re(w) > \alpha$.

By a Mellin inversion formula (or by a standard argument for the integral representation for factorial series), we have

$$f(z) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} f^*(w)(1-z)^{-w}\,\mathrm{d}w \qquad (\sigma > \alpha). \tag{4.11}$$

Substituting this into (4.3) and using (4.4), we obtain

$$\begin{aligned} &\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} f^*(w)P_0(w)(1-z)^{-w}\,\mathrm{d}w \\ &\quad = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} f^*(w)\sum_{1\leq\ell\leq\mu} P_{\ell}(w)(1-z)^{-w+\ell}\,\mathrm{d}w + \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} g^*(w)(1-z)^{-w}\,\mathrm{d}w. \end{aligned}$$

By absolute convergence and by the changes of variables $w \mapsto w+j$, we are led to the difference equation

$$f^*(w) = \sum_{1\le \ell\le \mu} \frac{P_\ell(w+\ell)}{P_0(w)}\, f^*(w+\ell) + \frac{g^*(w)}{P_0(w)} \tag{4.12}$$

for $\Re(w) > \alpha$, with the additional property that $f^*(w) \to 0$ as $|w| \to \infty$ in the half-plane $|\arg(w)| < \pi$. But the right-hand side of (4.12) also gives a meromorphic continuation of $f^*(w)$ to the whole plane (up to the zeros of $P_0(w+\ell)$ for $\ell \ge 0$).

In particular, since α is a simple zero of $P_0(w)$ (see [9]),

$$f^*(w) \sim \frac{K'}{w-\alpha} \qquad (w \sim \alpha), \tag{4.13}$$

where the residue K' can be computed by

$$\begin{aligned} K' &:= \lim_{w\to\alpha} (w-\alpha) f^*(w) \\ &= \frac{1}{P_0'(\alpha)} \left(\sum_{1\le \ell\le \mu} P_\ell(\alpha+\ell) f^*(\alpha+\ell) + g^*(\alpha) \right). \end{aligned} \tag{4.14}$$

Mellin inversion and singularity analysis. Heuristically, the series form (4.6) results easily from (4.13) and the Mellin inversion formula (4.11) as follows. First, by (4.11) using (4.13), we expect that

$$f(z) \sim K'(1-z)^{-\alpha},$$

by formally shifting the integration line to $\Re(w) = \alpha - \varepsilon$ and by taking into account the residue of the integrand at $w = \alpha$.

Then by a formal application of the singularity analysis, we anticipate the approximation

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

so that K' should equal K.

Justifying the above quick analysis requires an estimate of $|f^*(\sigma \pm iT)|$ for large T. On the other hand, because of the presence of the factor $(1-z)^{-w}$, which can be exponentially large when $z \in \mathbb{C}$ lies near unity and when T grows, we need an estimate of $|f^*(\sigma \pm iT)|$ that decays at infinity at an exponential rate (in T). While this approach might be made rigorous, we prefer to develop another approach that is more general and does not rely on analytic properties but instead on algebraic manipulations of integrals.

Effective expressions for K. Iterating the right-hand side of (4.14) N times by the same difference equation (4.12), we deduce that

$$P_0'(\alpha)K = \sum_{1\le j\le (\mu-1)N+1} A_{N,j} f^*(\alpha+N+j-1) + K_N \tag{4.15}$$

for any $N \ge 1$, where $A_{1,j} := P_j(\alpha+j)$ for $1 \le j \le \mu$ and for $N \ge 2$,

$$A_{N,\ell} = \sum_{1\le i\le \mu} \frac{P_i(\alpha+N+\ell-1)}{P_0(\alpha+N+\ell-1-i)}\, A_{N-1,\ell+1-i} \qquad (1 \le \ell \le N(\mu-1)+1), \tag{4.16}$$

and

$$\begin{aligned} K_N &= K_{N-1} + \sum_{1\le j\le (N-1)(\mu-1)+1} \frac{g^*(\alpha+N+j-2)}{P_0(\alpha+N+j-2)}\, A_{N-1,j} \\ &= g^*(\alpha) + \sum_{1\le i\le N-1} \sum_{1\le j\le i(\mu-1)+1} \frac{g^*(\alpha+i+j-1)}{P_0(\alpha+i+j-1)}\, A_{i,j} \\ &= g^*(\alpha) + \sum_{1\le j\le (N-1)\mu} \frac{g^*(\alpha+j)}{P_0(\alpha+j)} \sum_{1\le i\le j} A_{i,j+1-i}. \end{aligned}$$

By induction and the recursive expression for P_j's, we have $P_j(x) \asymp x^k$ for large x. From this and the estimate (4.5), we obtain

$$K = \frac{1}{P_0'(\alpha)} \lim_{N\to\infty} K_N = \frac{1}{P_0'(\alpha)} \left(g^*(\alpha) + \sum_{j\ge 1} \frac{g^*(\alpha+j)}{P_0(\alpha+j)} \sum_{1\le i\le j} A_{i,j+1-i} \right).$$

Define

$$B_j := \frac{1}{P_0(\alpha+j)} \sum_{1\le i\le j} A_{i,j+1-i}.$$

Then B_j is easily seen to satisfy (4.7) by using (4.16).

5. A Cauchy–Euler approach. In this section we develop a different approach to proving (4.6). The reason we write (4.3) in the form of a Cauchy–Euler differential equation is that if

$$f(z) = (1-z)^{-\alpha} \xi(1-z) \tag{5.1}$$

for $z \in \mathbb{C}$ near unity, where $\xi(z)$ is analytic at the origin, then

$$\sum_{1\le \ell\le \mu} (1-z)^\ell P_\ell(\vartheta) f(z) = O\left(|1-z|^{-\alpha+1}\right), \tag{5.2}$$

and thus the dominant (asymptotic) solutions are expected to be determined by the left-hand side of (4.3), with the right-hand side behaving as if it is independent of f. We apply again the Frobenius method to prove (5.1). Once (5.1) is proved, (4.3) is then solved asymptotically by extending our approach of *iterative linear operators* developed in [10].

The growth order of f. By using the formula

$$\vartheta^j f(z) = \sum_{0\le \ell\le j} (-1)^{j+\ell} S(j,\ell) (1-z)^\ell f^{(\ell)}(z),$$

where the $S(j,\ell)$'s represent the Stirling numbers of the second kind, we can rewrite (4.3) in the form

$$\sum_{0\le j\le k} L_j(z) (1-z)^j f^{(j)}(z) = g(z)$$

for some polynomials L_j of degree k. In particular, $L_k(z) = z^\mu (1-z)^k$. Therefore, the possible singularities of (4.3) are the two zeros $z=0$ and $z=1$ of the polynomial

$z^\mu(1-z)^k$, $z=\infty$, and the singularities of $g(z)$ ($z=1$ only). In particular, the singularity $z=1$ is a *regular singular point*.

Before applying the Frobenius method (see [24]), we use suitable operators to convert the nonhomogeneous equation into a homogeneous one. We start from the relation

$$(\vartheta+j)(1-z)^j=0 \qquad (j\in\mathbb{Z}).$$

Then we can annihilate the nonhomogeneous term in (4.3) by multiplying both sides by a sufficient number of the above annihilating operators, giving

$$\left(\prod_{-1\le j<k+\mu}(\vartheta+j)\right)\left(P_0(\vartheta)f(z)-\sum_{1\le\ell\le\mu}(1-z)^\ell P_\ell(\vartheta)f(z)\right)=0. \tag{5.3}$$

This means that if $f(z)$ solves the nonhomogeneous equation (4.3), then it is also a solution of (5.3) since the dominant growth order of f near the singularity $z=1$ is determined by the zero of the new indicial equation

$$\left(\prod_{-1\le j\le d}(\vartheta+j)\right)P_0(\vartheta)$$

with the largest real part. By the Frobenius method (see [24]), we then deduce the required estimate (5.1), from which (5.2) follows.

Method of iterative linear operators. As in [10], let α_j, $1\le j\le k$, be the zeros of $P_0(z)$ arranged in decreasing order of their real parts (see [9])

$$\alpha=\alpha_1>\Re(\alpha_2)\ge\cdots\ge\Re(\alpha_k).$$

Define

$$\mathbf{I}_\eta[f](z)=(1-z)^{-\eta}\int_0^z(1-x)^{\eta-1}f(x)\,\mathrm{d}x.$$

To solve (4.3), we first factor $P_0(\vartheta)$ into linear operators

$$P_0(\vartheta)f(z)=(\vartheta-\alpha_1)\cdots(\vartheta-\alpha_k)f(z)=R(z),$$

where

$$R(z):=g(z)+\sum_{1\le j\le\mu}(1-z)^jP_j(\vartheta)f(z),$$

and then solve the linear equations one after another, giving (see [10])

$$f(z)=(\mathbf{I}_{\alpha_k}\circ\mathbf{I}_{\alpha_{k-1}}\circ\cdots\circ\mathbf{I}_\alpha)[R](z).$$

By using successive integrations by parts (see [8]) or the inductive arguments used in [10], we deduce that

$$f(z)=\frac{1}{P_0'(\alpha)}\,\mathbf{I}_\alpha[R](z)+T_1(z),$$

where $[z^n]T_1(z) = o(n^{\alpha-1})$.

Then (see [10])

$$\begin{aligned} Q_n = [z^n]f(z) &= \frac{1}{P_0'(\alpha)}[z^n]\mathbf{I}_\alpha[R](z) + o(n^{\alpha-1}) \\ &= \frac{1}{P_0'(\alpha)} \sum_{0\le\ell<n} \frac{[z^\ell]R(z)}{\ell+1} \prod_{\ell+2\le j\le n} \left(1+\frac{\alpha-1}{j}\right) + o(n^{\alpha-1}) \\ &= \frac{K''}{P_0'(\alpha)\Gamma(\alpha)}\, n^{\alpha-1} + o(n^{\alpha-1}), \end{aligned}$$

where

$$K'' = \int_0^1 (1-x)^{\alpha-1} R(x)\,\mathrm{d}x = \sum_{\ell\ge 0} \frac{\ell![z^\ell]R(z)}{\alpha(\alpha+1)\cdots(\alpha+\ell)}.$$

Iteration of the leading constant. We next "mimic" the operations of the Mellin approach and show that

$$K'' = P_0'(\alpha)K,$$

and this will complete the proof of Theorem 4.1.

LEMMA 5.1. *Let $\omega(x)$ and $\beta(x)$ be two polynomials of degree at most k. Let $h(x)$ be defined in the unit interval. If h satisfies*

(i) *$h^{(j)}(0) = 0$ for $0 \le j < k$, and*

(ii) *the integral*

$$h^*(w) := \int_0^1 (1-x)^{\eta-1} h(x)\,\mathrm{d}x$$

converges,

then for any $\eta \in \mathbb{C}$ for which $\beta(\eta) \neq 0$,

$$\int_0^1 (1-x)^{\eta-1}\omega(\vartheta_x)\beta(\vartheta_x)^{-1}h(x)\,\mathrm{d}x = \frac{\omega(\eta)}{\beta(\eta)}\, h^*(\eta), \tag{5.4}$$

where $\beta(\vartheta_x)^{-1}$ represents the inverse operator of the differential operator $\beta(\vartheta_x)$ and $\vartheta_x := (1-x)(\mathrm{d}/\mathrm{d}x)$.

Proof. Obviously, (5.4) holds when $\omega(x) = \beta(x)$. It is also easily checked, by integration by parts, for the two fundamental cases $\omega(x) = x$, $\beta(x) = 1$ and $\omega(x) = 1$, $\beta(x) = x - \nu$ for $\nu \neq \eta$.

When $\omega(x)$ is a polynomial of degree at most k, we first write $\omega(x) = (x-\nu)\omega_1(x)$, where ν is a zero of $\omega(x)$. Then

$$\begin{aligned} \int_0^1 (1-x)^{\eta-1}\omega(\vartheta_x)h(x)\,\mathrm{d}x &= \int_0^1 (1-x)^{\eta-1}(\vartheta_x-\nu)\omega_1(\vartheta_x)h(x)\,\mathrm{d}x \\ &= (\eta-\nu)\omega_1(\vartheta_x)h^*(\eta). \end{aligned}$$

Repeating the same argument, we have

$$\int_0^1 (1-x)^{\eta-1}\omega(\vartheta_x)h(x)\,\mathrm{d}x = \omega(\eta)h^*(\eta)$$

and

$$\int_0^1 (1-x)^{\eta-1}\beta(\vartheta_x)^{-1}h(x)\,\mathrm{d}x = \frac{h^*(\eta)}{\beta(\eta)}.$$

Similarly, we derive (5.4). □

Define

$$\Lambda_f(\alpha) := \int_0^1 (1-x)^{\alpha-1}P_0(\vartheta_x)f(x)\,\mathrm{d}x.$$

Note that $\Lambda_f(\alpha) = K''$. By substituting (4.3) into the above integral and by applying (5.4), we have

$$\Lambda_f(\alpha) = g^*(\alpha) + \sum_{1\le\ell\le\mu} \frac{P_\ell(\alpha+\ell)}{P_0(\alpha+\ell)}\,\Lambda_f(\alpha+\ell), \tag{5.5}$$

where $g^*(\alpha)$ is defined as in (4.4). By definition, we have

$$K'' = g^*(\alpha) + \sum_{1\le\ell\le\mu} \frac{P_\ell(\alpha+\ell)}{P_0(\alpha+\ell)}\,\Lambda_f(\alpha+j).$$

It follows by (4.12) and the property $\Lambda_f(x) \to 0$ as $x \to \infty$ that $K'' = P_0'(\alpha)K$, as required.

More refined approximations. We can refine the above analysis and derive the effective approximation

$$Q_n = \frac{K}{\Gamma(\alpha)}\,n^{\alpha-1} - \frac{1}{2^k-2^s}\sum_{1\le j\le k} 2^{j-1+q_j+\cdots+q_k} + O\left(n^{\alpha-2} + n^{\Re(\alpha_2)-1}(\log n)^{\kappa-1}\right), \tag{5.6}$$

where κ denotes the largest multiplicity of zeros with real parts equal to $\Re(\alpha_2)$ (see [10]).

Hypergeometric cases: $\mu = 1$. When the query pattern is of the form (2.15), we can obtain more explicit expressions. In this case, $\mu = 1$ and $P_1(x) = x^k - 2^k$. Then

$$K = \frac{1}{P_0'(\alpha)}\sum_{j\ge0} g^*(\alpha+j)\prod_{1\le\ell\le j}\frac{P_1(\alpha+\ell)}{P_0(\alpha+\ell)},$$

where, by a lengthy calculation,

$$g^*(w) := \frac{(s+1)2^k-2^s}{w-1} + \frac{2^k\sigma_0 - 2^{k+1} + 2^{k-s}+1}{w} - \sum_{1\le\ell\le k}\frac{P_0(-\ell)\sigma_\ell - P_1(1-\ell)\sigma_{\ell-1}}{w+\ell},$$

with $\sigma_\ell := (-1)^\ell \sum_{\ell\le j<k}\binom{j}{\ell}Q_j$. From this expression of g, we can further rewrite the above series form for K in terms of generalized hypergeometric functions.

$\mu \ge 2$. When $\mu \ge 2$, the solution to the recurrence (4.7) is, in general, less explicit. But there are special cases when $\mu > 1$ can be reduced to $\mu = 1$. These occur when the query patterns are of the form

$$[\underbrace{0,\dots,0}_{k-\mu-s+1},\underbrace{1,\dots,1}_{s},0,\dots,0] \qquad (1\le s\le k-\mu+1),$$

TABLE 5.1

Approximate numerical values of α and $Q_n/n^{\alpha-1}$ for all generating query patterns (up to cyclic rotations) for $k = 3$, 4, 5, and 6. Note that the periodic patterns $[0,1,0,1]$, $[0,1,0,1,0,1]$, $[0,0,1,0,0,1]$ and $[0,1,1,0,1,1]$ are not shown. The total number of different patterns (up to rotations) needed to cover all 2^k possible query forms is given by (2.4).

k	Query	$\alpha \approx$	$Q_n/n^{\alpha-1} \approx$
3	[0, 0, 1]	1.71618 86589 93105	1.88700 34494 16788
	[0, 1, 1]	1.39485 86738 66065	4.30626 60684 05608
4	[0, 0, 0, 1]	1.78995 09772 69481	1.66842 10542 80183
	[0, 0, 1, 1]	1.56155 28128 08830	2.91912 77264 05377
	[0, 1, 1, 1]	1.30555 31614 36616	6.16148 43696 18751
5	[0, 0, 0, 0, 1]	1.83323 02942 30338	1.55981 70883 35171
	[0, 0, 0, 1, 1]	1.65556 26632 48591	2.46232 74709 13660
	[0, 0, 1, 0, 1]	1.65556 26632 48591	2.21247 44030 89626
	[0, 0, 1, 1, 1]	1.46323 88095 76994	4.05206 48849 27873
	[0, 1, 1, 0, 1]	1.46323 88095 76994	3.24485 54614 39396
	[0, 1, 1, 1, 1]	1.24956 22677 97953	8.07657 98073 98937
6	[0, 0, 0, 0, 0, 1]	1.86170 55962 67907	1.49483 95666 52432
	[0, 0, 0, 0, 1, 1]	1.71618 86589 93105	2.23506 40352 34654
	[0, 0, 0, 1, 0, 1]	1.71618 86589 93105	2.04449 79159 05194
	[0, 0, 0, 1, 1, 1]	1.56155 28128 08830	3.35792 47385 71823
	[0, 0, 1, 1, 0, 1]	1.56155 28128 08830	2.79761 13698 36274
	[0, 0, 1, 0, 1, 1]	1.56155 28128 08830	3.03451 26969 10640
	[0, 1, 1, 1, 0, 1]	1.39485 86738 66065	3.95385 24193 61293
	[0, 0, 1, 1, 1, 1]	1.39485 86738 66065	5.25849 03652 85129
	[0, 1, 1, 1, 1, 1]	1.21106 87077 39977	10.0301 95663 53780

TABLE 5.2

Coefficients of $(w+j)^{-1}$, $-1 \le j < k+\mu$ in $g^*(w)$ for $k \le 5$.

Query	-1	0	1	2	3	4	5	6	7
$[0,0,1]$	14	13	-64	69	-32				
$[0,1,1]$	20	11	-64	65	-32				
$[0,0,0,1]$	30	73	-320	298	114	-195			
$[0,0,1,1]$	44	69	-320	342	60	-195			
$[0,1,1,1]$	56	$\frac{571}{9}$	$-\frac{2752}{9}$	$\frac{1034}{3}$	$\frac{200}{9}$	$-\frac{1625}{9}$			
$[0,0,0,0,1]$	62	273	-1280	2670	-5926	8425	-4224		
$[0,0,0,1,1]$	92	265	-1280	2390	-4900	7657	-4224		
$[0,0,1,0,1]$	76	$\frac{994}{9}$	$-\frac{15400}{9}$	$\frac{46513}{9}$	$-\frac{33088}{3}$	$\frac{180400}{9}$	$-\frac{219772}{9}$	$\frac{142805}{9}$	$-\frac{12320}{3}$
$[0,0,1,1,1]$	120	$\frac{2317}{9}$	$-\frac{11360}{9}$	$\frac{6634}{3}$	$-\frac{37144}{9}$	$\frac{62165}{9}$	$-\frac{12320}{3}$		
$[0,1,1,0,1]$	88	$\frac{524}{9}$	$-\frac{4616}{3}$	$\frac{13757}{3}$	$-\frac{82924}{9}$	$\frac{49202}{3}$	$-\frac{61148}{3}$	$\frac{122495}{9}$	-3608
$[0,1,1,1,1]$	144	$\frac{2099}{9}$	$-\frac{10424}{9}$	$\frac{5818}{3}$	$-\frac{29600}{9}$	$\frac{51647}{9}$	-3608		

the expected cost of which can be computed via that of the special patterns (2.15) and the relations (2.3).

Tables. For concreteness, we give numerical approximations to $K/\Gamma(\alpha)$ in Table 5.1 and the expansions of $g^*(w)$ for $k \le 5$ in Table 5.2.

Note that the general terms in (4.6) converge slowly; thus it is more efficient, for numerical purposes, to use the expression (4.15); see Table 5.1. The idea is roughly to take a sufficiently large N, say 1000, to compute the first N terms as precisely as we can, and then to estimate the errors by the values of $f^*(\alpha+N)$, which can be easily computed by

$$f^*(\alpha+N)=\sum_{j\ge k}\frac{Q_j j!}{(\alpha+N)\cdots(\alpha+N+j)},$$

the series itself being an asymptotic expansion.

6. *k*-d-*t* trees. In this section we extend our approach to k-d-t trees, $t\in\mathbb{N}$, which are locally balanced versions of k-d trees (in which all trees of sizes larger than $2t$ have both subtrees of sizes at least t); see [11]. For simplicity, trees of sizes $\le 2t$ are not rearranged.

The probabilistic model for partial match queries remains the same as above. For convenience, we use the same set of notations as for k-d trees (indexed by t when ambiguity may arise). Let $Q_{n,t}=Q_n[q_1,\dots,q_k;t]$ stand for the expected number of nodes visited when performing the range search algorithm of a random partial match query in a random k-d-t tree of n nodes. Then $Q_{n,0}=Q_n$. Cunto, Lau, and Flajolet [11] showed that

$$Q_{n,t}\sim C_t n^{\alpha-1}$$

for some constant C_t, where $\alpha>1$ solves the equation $P_0(\alpha)=0$, with

$$P_0(x):=\left((x+t)^{\overline{t+1}}\right)^{k-s}\left((x+t+1)^{\overline{t+1}}\right)^{s}-\left((t+2)^{\overline{t+1}}\right)^{k}. \tag{6.1}$$

We show that C_t can be computed as above but with more complicated components.

LEMMA 6.1 (basic recurrences). *Define* $\varpi_{n,j}:=\binom{j}{t}\binom{n-1-j}{t}/\binom{n}{2t+1}$. *The expected search cost* $Q_{n,t}$ *for fixed* $t\in\mathbb{N}$ *satisfies the system of recurrences*

$$\begin{cases} Q_n[q_1,\dots,q_k;t]=1+2\displaystyle\sum_{1\le j<n}\rho_{n,j}(q_1)\varpi_{n,j}Q_j[q_2,\dots,q_k,q_1;t],\\ \qquad\qquad\vdots\\ Q_n[q_k,q_1,\dots,q_{k-1};t]=1+2\displaystyle\sum_{1\le j<n}\rho_{n,j}(q_k)\varpi_{n,j}Q_j[q_1,\dots,q_k;t] \end{cases} \tag{6.2}$$

for $n\ge 2t+1$, *with the initial values* $Q_{n,t}\equiv Q_n$ *for* $n\le 2t$.

Differential equations. Define the differential operator

$$J_q[\varphi](z)=\left((\vartheta+t)\cdots(\vartheta+2t)+\frac{q(t+1)}{z}(\vartheta+t+1)\cdots(\vartheta+2t)\right)\varphi(z)\qquad(q\in\{0,1\}).$$

From (6.2), we derive a system of differential equations satisfied by the generating functions

$$f_j^{[t]}(z):=Q[q_j,\dots,q_k,q_1,\dots,q_{j-1};t](z):=\sum_{n\ge 1}Q_{n,t}z^n.$$

LEMMA 6.2. *Let* $\lambda_t := (2t+2)!/(t+1)!$. *For any query pattern* $[q_1, \dots, q_k]$,

$$\begin{cases} J_{q_1}[f_1^{[t]}] = J_{q_1}\left[\dfrac{z}{1-z}\right] + \lambda_t f_2^{[t]}, \\ \qquad\vdots \\ J_{q_k}[f_k^{[t]}] = J_{q_k}\left[\dfrac{z}{1-z}\right] + \lambda_t f_1^{[t]}. \end{cases}$$

Proof (sketch). When $q_j = 0$, we have

$$\vartheta^{\overline{2t+1}}\left(f_j^{[t]} - \frac{z}{1-z}\right) = \lambda_t \vartheta^{\overline{t}} f_{j+1}^{[t]} \qquad (1 \le j \le k),$$

where $f_{k+1}^{[t]} := f_1^{[t]}$.

When $q_j = 1$, the differential equation becomes

$$\vartheta^{\overline{2t+2}}\left(z f_j^{[t]} - \frac{z^2}{1-z}\right) = \lambda_t \vartheta^{\overline{t+1}}(z f_{j+1}^{[t]}).$$

By applying the identity $\vartheta\left((1-z)\varphi(z)\right) = (1-z)(\vartheta-1)\varphi(z)$, we obtain

$$\left((\vartheta+t)\cdots(\vartheta+2t) + \frac{(t+1)}{z}(\vartheta+t+1)\cdots(\vartheta+2t)\right) f_j^{[t]} = \lambda_t f_{j+1}^{[t]}. \qquad \square$$

COROLLARY 6.3.

$$\left(J_{q_k} \circ \cdots \circ J_{q_1}\right)[f_1^{[t]}] - \lambda_t^k f_1^{[t]} = \sum_{1 \le j \le k} \lambda_t^{j-1} \left(J_{q_k} \circ \cdots \circ J_{q_j}\right)\left[\frac{z}{1-z}\right].$$

Differential equations: Normal form. We multiply, as in the case of k-d trees, both sides by $z^{\mu_{\mathbf{q}}(t)}$, where $\mu_{\mathbf{q}}(t)$ denotes the "minimum exponent"

$$\mu_{\mathbf{q}}(t) := (k-v)(t+1) + 1,$$

with v being the last position of the first block of 1's (from left to right), and $\mu_{\mathbf{q}}(t) := 0$ if $s = 0$.

Then the generating function

$$f(z) := f_1^{[t]}(z) - \sum_{1 \le j < k(t+1)} Q_j z^j$$

satisfies the differential equation

$$P_0(\vartheta) f = \sum_{1 \le \ell \le \mu_{\mathbf{q}}(t)} (1-z)^\ell P_\ell(\vartheta) f + g,$$

where the P_ℓ's are polynomials of degree $k(t+1)$ and

$$g(z) = z^{\mu_{\mathbf{q}}(t)} \sum_{1 \le j \le k} \lambda_t^{j-1} \left(J_k \circ \cdots \circ J_j\right)\left[\frac{z}{1-z}\right](z) + z^{\mu_{\mathbf{q}}(t)} \lambda_t^k \sum_{1 \le j < k(t+1)} Q_j z^j$$
$$- z^{\mu_{\mathbf{q}}(t)} \left(J_k \circ \cdots \circ J_1\right)\left[\sum_{1 \le j < k(t+1)} Q_j z^j\right](z).$$

In particular, $P_0(x)$ is as given in (6.1).

A series form for K_t. Following the same method of proof as for k-d trees, we can derive a series representation for K_t.

THEOREM 6.4. *The expected cost $Q_{n,t}$ satisfies the approximation*

$$Q_{n,t} \sim \frac{K_t}{\Gamma(\alpha)}\, n^{\alpha-1},$$

where

$$K_t = \frac{1}{P_0'(\alpha)} \sum_{j\ge 0} g^*(\alpha+j) B_j,$$

with the series being absolutely convergent. Here $g^*(w) := \int_0^1 g(x)(1-x)^{w-1}\,\mathrm{d}x$ and B_j is defined recursively by

$$B_j = \sum_{1\le \ell\le \mu_{\mathbf{q}}(t)} \frac{P_\ell(\alpha+j)}{P_0(\alpha+j)}\, B_{j-\ell} \qquad (j\ge 1),$$

with the initial values $B_0 = 1$ and $B_j = 0$ for $j < 0$.

The method of proof is the same as that for k-d trees. In particular, the estimates $g^*(w) = O(|w|^{-\mu_{\mathbf{q}}(t)-1})$ and $B_j = O(|j|^{\mu_{\mathbf{q}}(t)-q_k-1})$ hold. Note that the generating function $B(z) := \sum_j B_j z^j$ satisfies the differential equation

$$\left(\prod_{t<\ell\le 2t}\left(\ell+\alpha-\frac{\mu_{\mathbf{q}}(t)z}{1-z}+z\mathbb{D}_z\right)\right)\times\left(\frac{q_1(t+1)}{1-z}+t+\alpha-\frac{\mu_{\mathbf{q}}(t)z}{1-z}+z\mathbb{D}_z\right)\times\cdots$$

$$\times\left(\prod_{t<\ell\le 2t}\left(\ell+\alpha-\frac{\mu_{\mathbf{q}}(t)z}{1-z}+z\mathbb{D}_z\right)\right)\left(\frac{q_k(t+1)}{1-z}+t+\alpha-\frac{\mu_{\mathbf{q}}(t)z}{1-z}+z\mathbb{D}_z\right)B(z)=0,$$

and the indicial equation is given by

$$\prod_{1\le j\le k}\left((-\zeta-\mu+(q_{k-j+1}+j-1)(t+1))\prod_{1\le\ell\le t}(-\zeta-\mu+(j-1)(t+1)+\ell)\right),$$

with the dominant (simple) zero being $\zeta = -\mu + q_k$.

All coordinates specified: $s = k$. In this case, the two Q_n's on both sides of (6.2) are the same; thus the k recurrences reduce to a single one. The associated differential equation is

$$\left(\vartheta^{\overline{2t+2}} - \lambda_t \vartheta^{\overline{t+1}}\right)(zf(z)) = \frac{(2t+2)!}{1-z},$$

with suitable initial conditions. Note that the polynomial $(x+t+1)\cdots(x+2t+1)-\lambda_t$ has 1 as the simple, largest zero (in real part). By applying [10, Theorem 1], we have

$$Q_n[1,\dots,1;t] \sim \frac{\log n}{H_{2t+2}-H_{t+1}},$$

another well-known result; see [11, 30, 36].

Hypergeometric cases: $\mu_{\mathbf{q}}(t) = 1$. For query patterns of the form (2.15), we have $\mu_{\mathbf{q}}(t) = 1$, and

$$P_1(x) = \prod_{t\le j\le 2t}(x+j)^k - \lambda_t^k.$$

7. Conclusions. We derived an effective expression for the leading constant of the expected cost used for random partial match queries in random k-d trees. The proposed approach for solving linear differential equations with polynomial coefficients is very general, and its success relies on the close interplay between complex-analytic methods and elementary methods (without complex analysis).

Complex-analytic methods, when they apply, usually give efficient approximations and neat expressions, but at the cost that stronger analytic properties are needed. On the other hand, elementary methods are computationally less efficient but can provide more general results under weaker conditions. A combination of both (heuristically or rigorously) usually yields very powerful tools, as already shown by the problems studied in [7, 8, 9, 10] and in this paper.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1972.

[2] J. L. Bentley, *Multidimensional binary search trees used for associative searching*, Comm. ACM, 18 (1975), pp. 509–517.

[3] J. L. Bentley, *K-d trees for semidynamic point sets*, in Proceedings of the Sixth Annual Symposium on Computational Geometry, ACM, New York, 1990, pp. 187–197.

[4] J. L. Bentley and J. H. Friedman, *Data structures for range searching*, ACM Comput. Surveys, 11 (1979), pp. 397–409.

[5] E. Bertino, B. Catania, B. C. Ooi, R. Sacks-Davis, K. L. Tan, J. Zobel, B. Shidlovsky, and B. Catania, *Indexing Techniques for Advanced Database Systems*, Kluwer Academic Publishers, Boston, 1997.

[6] P. Chanzy, L. Devroye, and C. Zamora-Cura, *Analysis of range search for random k-d trees*, Acta Inform., 37 (2001), pp. 355–383.

[7] H.-H. Chern, M. Fuchs, and H.-K. Hwang, *Phase changes in random point quadtrees*, ACM Trans. Algorithms, to appear.

[8] H.-H. Chern and H.-K. Hwang, *Phase changes in random m-ary search trees and generalized quicksort*, Random Structures Algorithms, 19 (2001), pp. 316–358.

[9] H.-H. Chern and H.-K. Hwang, *Partial match queries in random quadtrees*, SIAM J. Comput., 32 (2003), pp. 904–915.

[10] H.-H. Chern, H.-K. Hwang, and T.-H. Tsai, *An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms*, J. Algorithms, 44 (2002), pp. 177–225.

[11] W. Cunto, G. Lau, and P. Flajolet, *Analysis of kdt-trees: kd-trees improved by local reogranisations*, in Algorithms and Data Structures, Lecture Notes in Comput. Sci. 382, Springer, Berlin, 1989, pp. 24–38.

[12] L. Devroye, J. Jabbour, and C. Zamora-Cura, *Squarish k-d trees*, SIAM J. Comput., 30 (2000), pp. 1678–1700.

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

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

[15] P. Flajolet, X. Gourdon, and P. Dumas, *Mellin transforms and asymptotics: Harmonic sums*, Theoret. Comput. Sci., 144 (1995), pp. 3–58.

[16] P. Flajolet and A. M. Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math., 3 (1990), pp. 216–240.

[17] P. Flajolet and P. Puech, *Partial match retrieval of multi-dimensional data*, J. ACM, 32 (1986), pp. 371–407.

[18] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, in preparation. Preliminary version available online at http://algo.inria.fr/flajolet/Publications/publist.html.

[19] J. H. Friedman, J. L. Bentley, and R. A. Finkel, *An algorithm for finding best matches in logarithmic expected time*, ACM Trans. Math. Software, 3 (1977), pp. 209–226.

[20] G. H. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, Addison-Wesley, Reading, MA, 1991.

[21] A. G. Gray and A. W. Moore, *"N-body" problems in statistical learning*, in Advances in Neural Information Processing Systems, T. K. Leen and T. G. Dietterich, eds., MIT Press, Cambridge, MA, 2001, pp. 521–527.

[22] P. J. Grother, G. T. Candela, and J. L. Blue, *Fast implementations of nearest neighbor classifiers*, Pattern Recognition, 30 (1997), pp. 459–465.
[23] T. N. Hibbard, *Some combinatorial properties of certain trees with applications to searching and sorting*, J. ACM, 9 (1962), pp. 13–28.
[24] E. L. Ince, *Ordinary Differential Equations*, Dover, New York, 1926.
[25] D. T. Lee and C. K. Wong, *Worst-case analysis for region and partial region searches in multidimensional binary search trees and quad trees*, Acta Inform., 9 (1977), pp. 23–29.
[26] C. Martínez, A. Panholzer, and H. Prodinger, *Partial match queries in relaxed multidimensional search trees*, Algorithmica, 29 (2001), pp. 181–204.
[27] R. Neininger, *Asymptotic distributions for partial match queries in K-d trees*, Random Structures Algorithms, 17 (2000), pp. 403–427.
[28] R. C. Nichol, S. Chong, A. J. Connolly, S. Davies, C. Genovese, A. M. Hopkins, C. J. Miller, A. W. Moore, D. Pelleg, G. T. Richards, J. Schneider, I. Szapudi, and L. Wasserman, *Computational AstroStatistics: Fast and Efficient Tools for Analysing Huge Astronomical Data Sources*, preprint, 2002. Available online at http://arxiv.org/abs/astro-ph/0110230.
[29] S. M. Omohundro, *Efficient algorithms with neural network behavior*, Complex Systems, 1 (1987), pp. 273–347.
[30] P. V. Poblete and J. I. Munro, *The analysis of a fringe heuristic for binary search trees*, J. Algorithms, 6 (1985), pp. 336–350.
[31] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*, Springer-Verlag, New York, 1985.
[32] J. Reiss, J.-J. Aucouturier, and M. Sandler, *Efficient multidimensional searching routines for music information retrieval: Use k-d trees to retrieve music information*, in Proceedings of the Second Annual International Symposium on Music Information Retrieval (Bloomington, IN), J. S. Downie and D. Bainbridge, eds., 2001, pp. 163–171.
[33] H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, MA, 1990.
[34] H. Samet, *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*, Addison-Wesley, Reading, MA, 1990.
[35] F. Schwarzer and I. Lotan, *Approximation of protein structure for fast similarity measures*, in Proceedings of the Seventh Annual ACM International Conference on Computational Biology (RECOMB), ACM, New York, 2003, pp. 267–276.
[36] R. Sedgewick, *Quicksort*, Garland Publishing, New York, 1980.
[37] S. S. Skiena, *Who is interested in algorithms and why? Lessons from the Stony Brook algorithms*, ACM SIGACT News, 30 (1999), pp. 65–74. Available online at http://www.cs.sunysb.edu/~skiena/hits.