Partial Match Retrieval of Multidimensional Data

PHILIPPE FLAJOLET

INRIA, Rocquencourt, France

AND

CLAUDE PUECH

Université de Paris-Sud, Orsay, France and Ecole Normale Supérieure, Montrouge, France

Abstract. A precise analysis of partial match retrieval of multidimensional data is presented. The structures considered here are multidimensional search trees (k-d-trees) and digital tries (k-d-tries), as well as structures designed for efficient retrieval of information stored on external devices. The methods used include a detailed study of a differential system around a regular singular point in conjunction with suitable contour integration techniques for the analysis of k-d-trees, and properties of the Mellin integral transform for k-d-tries and extendible cell algorithms.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*sorting and searching*; G.2.1 [**Discrete Mathematics**]: Combinatorics—*counting problems*; *generating functions*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*search process*

General Terms: Algorithms, Performance

Additional Key Words and Phrases: Analysis of algorithms, data structures, multidimensional search, partial match, trees

1. *Introduction*

Methods for *retrieval of multidimensional data* are of prime importance to the design of database systems and to specific applications, including the management of geographical data or graphics algorithms. The ancestry of most currently developed algorithms is to be found in early works by Rivest [20], where *hashing* and *digital techniques* are explored, and by Bentley [1] and Finkel and Bentley [6], who proposed *quadtrees* and *k-d-trees*, which are *comparison-based structures.* A description of early algorithms appears in Section 6.5 of Knuth's book [13]. Recent developments in the context of large external files combine some of these techniques with ideas derived from *dynamic hashing schemes* for single-attribute records (virtual hashing [15], dynamic hashing [14], extendible hashing [5]); a few such

Some of this work was started while P. Flajolet was visiting the Tata Institute of Fundamental Research in Bombay.

Authors' addresses: P. Flajolet, INRIA, Domaine de Voluceau—Rocquencourt, B.P. 105, 78153 Le Chesnay Cedex, France; C. Puech, Laboratoire de Recherche en Informatique, UA 410 "Al Khowarizmi" du CNRS, Université de Paris-Sud, Bât. 490, 91405 Orsay, France, and Ecole Normale Supérieure, 1 Rue Maurice Arnoux, 92120 Montrouge, France.

examples are the *grid-file* of Nievergelt et al. [17], the *extendible* cell method [21], and the multidimensional extendible hashing algorithm of [16].

This paper describes evaluation methods for the major multidimensional search algorithms. We concentrate here on the problem known as *partial match retrieval*, where all records in a file having specified values for some of their attributes are to be found. Contrary to the case of single attribute search, no general algorithm is known for locating a record in a file of size n using $O(\log n)$ time and linear storage. (It is indeed conjectured that no such algorithm exists; see [20].) We prove here that the average search cost in a file of size n, containing k-dimensional records, when s attributes are specified ($0 < s < k$) is

(a) *For k-d-trees*: $O(n^{1-s/k+\theta(s/k)})$ field comparisons where $\theta(u)$ is a strictly positive function of u for $0 < u < 1$, with maximum value 0.07. This result is of interest since it disproves an (often quoted) old claim of Bentley [1] that k-d-trees perform in expected time $O(n^{1-s/k})$. Such a bound appears to hold only in the static case when the underlying tree structure is a perfect tree.

(b) *For k-d-tries*: $O(n^{1-s/k})$ bit comparisons where the implied constant in the $O(\ )$ is precisely characterized and turns out to be quite small. This result is a useful complement to some of Rivest's analyses [20] made under a different model, which suggested a higher order of $O(n^{\log_2(2-s/k)})$ for k-d-tries.

(c) *For grid-file algorithms*: $O(n^{1-s/k})$ page accesses; there again the implied constants can be precisely determined.

A comparison of these results shows that, for multidimensional search trees, digital methods asymptotically outperform comparison-based techniques. As an example, partial match retrieval of 2-dimensional records with one attribute specified has average cost:

$$O(n^{(\sqrt{17}-3)/2}) = O(n^{0.56}) \qquad \text{for 2-d-trees}$$

and

$$O(n^{1/2}) \qquad \text{for 2-d-tries.}$$

Performances of the type $O(n^{1-s/k})$ have been conjectured to be optimal by Rivest [20].

We feel that also of particular interest are the proof techniques employed in this paper (especially in case (a), for which previous analyses appear to be invalid):

(a) For k-d-trees, we start by setting up a system of *integral equations* for adequately chosen generating functions of costs. The system transforms into a *linear differential system* (with variable coefficients) of order $2k - s$, which does not seem to admit closed-form solutions. Indeed, the shape of our final results strongly suggests that no such form exists and that no elementary combinatorial approach is likely to be workable. We then proceed to study the way the system becomes singular, and with the help of classical results from the theory of "*regular singular points*" of differential systems, we obtain the asymptotic behavior of cost-generating functions around their common singularity. We then use the *Cauchy integral formula* for Taylor coefficients of power series in conjunction with suitable *contours of integration* (in a manner similar to that of [7]) to conclude the analysis of k-d-trees.

(b–c) There, we set up in each case a system of *difference equations* for generating functions of costs that can be solved explicitly. This leads to exact expressions for the average case behavior of algorithms considered. We then appeal to Mellin

transform techniques (see [13])[1] to derive the results stated in (b) and (c) relative to k-d-tries and grid-file algorithms.

It should be stressed that the methods used here are of a rather wide applicability: those of type (a) could serve to derive direct asymptotic evaluations for a number of comparison-based algorithms; methods of type (b–c) may be used to analyze in detail a number of data structures and algorithms closely related to tries, like the double-chained trees [2], multiattribute trees [12], and the like. These analyses will be given in a companion paper (see [8], for a preliminary report).

2. *General Setting*

We consider the problem of retrieving *multiattribute records* that belong to some k-dimensional domain

$$D = D_1 \times D_2 \times \cdots \times D_k.$$

A *file* F is any finite subset of D and the *size* of F, usually denoted by n, in the sequel, is the number of elements in F. Our interest is in data structures for performing *partial match retrieval*: Given F and a query $q = (q_1, q_2, \ldots, q_k)$,

$$q \in (D_1 \cup \{*\}) \times (D_2 \cup \{*\}) \times \cdots \times (D_k \cup \{*\}),$$

one is asked to find all records in F satisfying query q, that is, to determine the subset $q(F)$ of F of records $r = (r_1, r_2, \cdots, r_k)$ in F satisfying for all j, where $1 \leq j \leq k$, such that $q_j \neq *$,

$$r_j = q_j.$$

Thus, a query q = (TOTO, *, 39, 35,000, *) asks for all (five-dimensional) records whose first attribute is TOTO, third attribute 39, and fourth attribute 35,000; the second and fifth attributes are left unspecified. The *specification pattern* of a query q is a word u of length k over the alphabet $\{S, *\}$ where $u_j = S$ if q_j is specified and $u_j = *$ if q_j is left unspecified. In the above example, the specification pattern is thus $S*SS*$.

In the sequel, for the sake of unity, we assume that each of the attribute domains is assimilated to the real interval [0; 1]; this is practically justified when the binary encodings of attributes are sufficiently long strings. Our analyses are relative to the *uniform* probabilistic model, where we assume that attributes in either files or queries are uniformly and independently distributed over the interval. As is well known, in the case of comparison-based algorithms, this model is equivalent to the more general model where attributes are only assumed to be *independently drawn* from any continuous distribution over any interval, so that there the uniform model is general enough. In the case of digital techniques, the uniform model constitutes an excellent approximation to real situations when superposed hashing is used and, in other cases, an optimistic model of varying accuracy, depending upon the particular structure of the data manipulated. However, our analyses can be easily generalized to cover biased probabilities of occurrences of bits or characters in records, and the orders of magnitude of expected case complexities appear to be only very slightly affected by this change in the model. Thus, our general conclusions remain valid for a wide range of situations.

[1] See, in particular, the section on radix exchange sort [13, p. 131] where Knuth uses Mellin transform techniques under the name of "Gamma function method."

The general pattern of our analyses is as follows: We let $c_{u,n}$, with n an integer and $u = u_1 u_2 \ldots u_k$ a specification pattern, denote the *expected cost* of a query with specification pattern u in a file of size n. We then introduce some *generating function* $c_u(z)$ of the sequence $\{c_{u,n}\}_{n\geq 0}$. We find in each case (a), (b), (c) that there are two operators Φ_* and Φ_S such that

$$\text{System } \Sigma: \begin{cases} c_u(z) = \Phi_{u_1}(c_{u'}(z)), \\ c_{u'}(z) = \Phi_{u_2}(c_{u''}(z)), \\ \quad\vdots \\ c_{u^{(k-1)}}(z) = \Phi_{u_k}(c_u(z)), \end{cases}$$

where u', u'', u''', ... designate the patterns obtained by circularly shifting the letters of u to the left by 1, 2, 3, ... positions. The structure of system Σ reflects the cyclical changes of the partitioning attributes in the multidimensional trees.

For k-d-trees, Φ_* and Φ_S turn out to be integral operators; for the other cases, they are difference operators.

3. *Multidimensional Binary Search Trees*

Multidimensional binary search trees (or k-d-trees) are constructed by repeated insertions from the file to be represented. At the root of the tree, we use the first field of the record stored there as a discriminator; we choose to go right or left by comparing the first field of the record to be inserted with the first field of the root (going to the left if it is smaller, going to the right otherwise). At the second level of the tree, the second attribute serves to discriminate records and so on, attributes $1, 2, 3, \ldots, k$ being used cyclically as discriminators. From the definition, it follows that 1-d trees coincide with the usual binary search trees.

A partial match query proceeds along the tree, branching to one side if the corresponding field is specified by the query or proceeding along both subtrees if the field is unspecified.

From the definition also follows that a k-d-tree can be viewed as a recursive partitioning of the underlying space according to alternative dimensions. Figure 1 represents a tree constructed from a file of seven elements together with the associated partitioning of the plane.

The main theorem that we prove for k-d-trees is as follows:

THEOREM 1. *The average cost, measured by the number of internal nodes traversed, of a partial match query of specification pattern u in a k-d-tree constructed by random insertions from a file of size n satisfies*

$$c_{u,n} = \gamma_u n^{1-s/k+\theta(s/k)}[1 + o(1)],$$

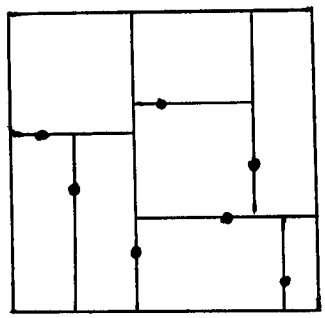
where γ_u is a strictly positive real constant[2] *and the function $\theta(x)$ is defined as the unique positive real root in the interval* [0; 1] *of the equation*

$$(\theta(x) + 3 - x)^x(\theta(x) + 2 - x)^{1-x} - 2 = 0,$$

so that, for $0 < x < 1$, $0 < \theta(x) < 0.07$.

3.1 BASIC EQUATIONS. Theorem 1 is proved through a chain of lemmas. Lemma 1 below expresses the recurrences satisfied by the quantities $c_{u,n}$, $c_{u',n}$, $c_{u'',n}$ with u, u', u'', ... being the successive left circular shifts of u. The natural expressions of these recurrences is in terms of corresponding generating functions.

[2] We discuss the problem of numerically estimating the constants γ_u in Section 5.



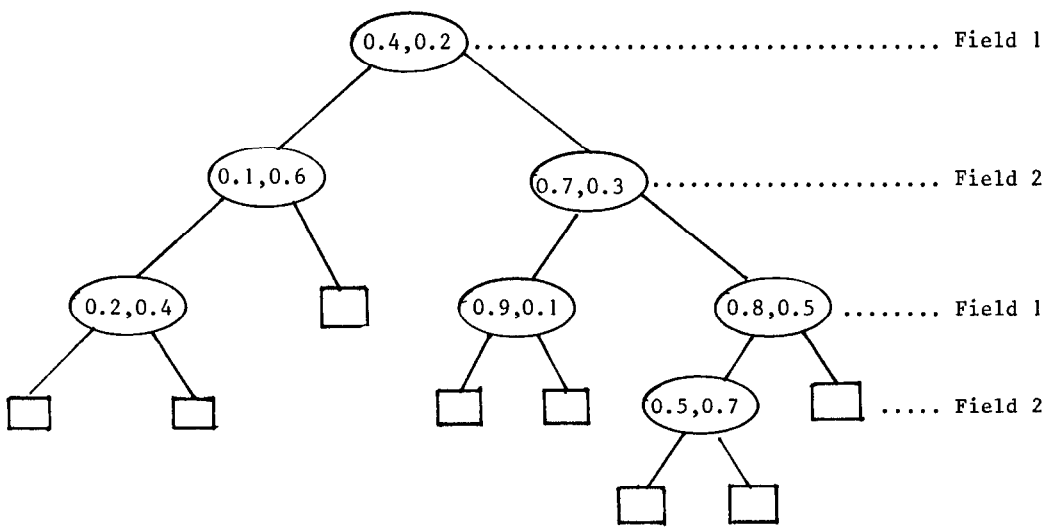


FIG. 1. The 2-d-tree associated with the file $F = \{(0.4, 0.2); (0.1, 0.6); (0.7, 0.3); (0.2, 0.4); (0.8, 0.5); (0.9, 0.1); (0.5, 0.7)\}$, elements arriving in the order in which they are listed, and a representation of the corresponding partitioning of $[0; 1] \times [0; 1]$.

LEMMA 1. *For each specification pattern u, define the generating functions*

$$c_u(z) = \sum_{n \geq 0} c_{u,n} z^n,$$

$$d_u(z) = \sum_{n \geq 0} c_{u,n}(n + 1) z^n.$$

One has

(*i*) *if* $u = *v$ (*i.e., the first attribute is unspecified*),

$$c_u(z) = \frac{1}{1 - z} - 1 + 2 \int_0^z c_{u'}(t) \frac{dt}{1 - t}.$$

(*ii*) *if* $u = Sv$ (*i.e., the first attribute is specified*),

$$d_u(z) = \frac{1}{(1 - z)^2} - 1 + 2 \int_0^z d_{u'}(t) \frac{dt}{1 - t}.$$

PROOF

(i) The average search cost in a fixed tree

$$t = t_1 \overset{\bigcirc}{\wedge} t_2$$

satisfies

$$c_u[t] = 1 + c_{u'}[t_1] + c_{u'}[t_2], \tag{1}$$

since, the first attribute being unspecified, one needs to visit the root of the tree and then recursively continue the search in both t_1 and t_2 with specification pattern u'. Taking expected values of (1), and noticing that the probability that t_1 contains p nodes, where, for any p, $0 \le p < n$, is uniformly $1/n$ (thus being independent of p), we find, for $n \ge 1$,

$$c_{u,n} = 1 + \frac{1}{n} \sum_{p=0}^{n-1} [c_{u',p} + c_{u',n-1-p}],$$

and by symmetry

$$c_{u,n} = 1 + \frac{2}{n} \sum_{p=0}^{n-1} c_{u',p}. \tag{2}$$

Taking corresponding generating functions and using (2) establish part (i) of the claim of the lemma.

(ii) The average search cost in a fixed tree

$$t = t_1 \overset{\bigcirc}{\wedge} t_2$$

when the first attribute is specified satisfies

$$c_u[t] = 1 + \frac{p+1}{n+1} c_{u'}[t_1] + \frac{n-p}{n+1} c_{u'}[t_2], \qquad \text{where} \quad p = |t_1|. \tag{3}$$

This corresponds to the fact that a search with first attribute specified proceeds along t_1 with probability $(p+1)/(n+1)$ and along t_2 with the complementary probability. Thus multiplying (3) by $(n+1)$, and taking average values over all possible trees t, we get in a similar manner, for $n \ge 1$,

$$\begin{aligned} (n+1)c_{u,n} &= (n+1) + \frac{1}{n} \sum_{p=0}^{n-1} [(p+1)c_{u',p} + (n-p)c_{u',n-1-p}], \\ (n+1)c_{u,n} &= (n+1) + \frac{2}{n} \sum_{p=0}^{n-1} (p+1)c_{u',p}. \end{aligned} \tag{4}$$

Part (ii) of the claim is nothing but the translation of recurrence (4) in terms of generating functions. □

We notice here that Lemma 1 uses an argument essentially equivalent to Bentley's observation [1] that the probability distribution of the shapes of k-d-trees constructed by n random insertions (forgetting about key values) coincides with the corresponding distribution on 1-d trees. This probability as a function of the shape of the tree is given in [13, sect. 6.2.2, ex. 5].

Our next step consists in reducing the equations of Lemma 1 to a vectorial differential system of order $2k - s$. The first k-components of the solution of the

system represent the quantities

$$d_u(z),\ d_{u'}(z),\ d_{u''}(z),\ \ldots,\ d_{u^{(k-1)}}(z), \tag{5}$$

and the remaining $k - s$ components are the primitives of those functions in (5) whose specification pattern starts with a star (*).

LEMMA 2. *The function $d_u(z)$ is the first component, $y_1(z)$, of the solution of the differential system of order $2k - s$:*

$$\frac{d}{dz}\,[\mathbf{y}(z)] = \Omega(z)\mathbf{y}(z) + \mathbf{b}(z), \tag{Σ}$$

where

$$\mathbf{y}(z) = (y_1(z),\ y_2(z),\ \ldots,\ y_{2k-s}(z))^T,$$
$$\mathbf{b}(z) = (b_1(z),\ b_2(z),\ \ldots,\ b_{2k-s}(z))^T,$$

with

$$b_i(z) = \frac{\epsilon_i}{(1-z)^3}$$

and

$$\epsilon_i = \begin{cases} 0 & \text{if } i > k, \\ 2 & \text{if } i \le k \text{ and } u_i = S, \\ 1 & \text{if } i \le k \text{ and } u_i = *. \end{cases}$$

The initial conditions are $y_j(0) = 0$. The transition matrix $\Omega(z)$ admits the block decomposition

$$\Omega(z) = \begin{pmatrix} A & C \\ B & D \end{pmatrix},$$

where matrices A, B, C, D have respective dimensions $k \times k$, $(k - s) \times k$, $k \times (k - s)$, $(k - s) \times (k - s)$ and elements given by

(*i*)
$$A_{ii} = \begin{cases} 0 & \text{if } u_i = S, \\ \dfrac{1}{z(1-z)} & \text{if } u_i = *; \end{cases}$$
$$A_{i,i+1 \bmod k} = \frac{2}{1-z}, \qquad \text{other elements are all zero;}$$

(*ii*)
$$B_{ij} = \begin{cases} 1 & \text{if } j \text{ is the rank of the } i\text{th unspecified attribute in } u, \\ 0 & \text{otherwise;} \end{cases}$$

(*iii*)
$$C = \frac{-1}{z^2(1-z)}\,B^T;$$

(*iv*) *D is the zero $(k - s) \times (k - s)$ matrix.*

PROOF. Let $\pi_1, \pi_2, \ldots, \pi_{k-s}$ be the ranks of the unspecified attributes in u; ranks are assumed to be numbered from 1. For instance, if $u = *SS**S*$, then $\pi_1 = 1$, $\pi_2 = 4$, $\pi_3 = 5$, $\pi_4 = 7$.

We set up a differential system for the quantities $y_1(z), y_2(z), \ldots, y_{2k-s}(z)$, where

$$y_j(z) = d_{u^{(j-1)}}(z) \qquad \text{for } j \text{ such that } 1 \le j \le k, \tag{6}$$

$$y_{k+j}(z) = \int_0^z d_{u^{(\pi_j-1)}}(t)\, dt \qquad \text{for } j \text{ such that } 1 \le j \le k-s. \tag{7}$$

The differential relations between the y_j's are obtained as follows:

(a) If $j \le k$ and $w = u^{(j-1)}$ starts with an S, differentiating the relation given by Lemma 1, part (ii), we have

$$d'_w(z) = \frac{2}{(1-z)^3} + \frac{2}{1-z}\, d_{w'}(z). \tag{8}$$

(b) If $j \le k$ and $w = u^{(j-1)}$ starts with a $*$, differentiating the relation given by Lemma 1(i), we find

$$c'_w(z) = \frac{1}{(1-z)^2} + \frac{2}{1-z}\, c_{w'}(z).$$

Multiplying this relation by z and adding to both members $c_w(z)$, we find

$$d_w(z) \equiv z c'_w(z) + c_w(z) = \frac{z}{(1-z)^2} + c_w(z) + \frac{2z}{1-z}\, c_{w'}(z).$$

We now multiply this last relation by $(1-z)$, differentiate and then multiply again by $1/(1-z)$, and isolate $d'_w(z)$, so that we get

$$d'_w(z) = \frac{1}{(1-z)^3} + \frac{1}{z(1-z)}\, d_w(z) - \frac{1}{z^2(1-z)} \int_0^z d_w(t)\, dt + \frac{2}{1-z}\, d_{w'}(z), \tag{9}$$

since

$$c_w(z) = \frac{1}{z} \int_0^z d_w(t)\, dt.$$

(c) Finally relation (7) is clearly equivalent to

$$y'_{k+j}(z) = y_{\pi_j}(z). \tag{10}$$

Putting together relations (8)–(10) leads to a differential system for the y_j's defined by (6) and (7) and the matrix form of this system is none other than the one given by the statement of the lemma. □

As an illustration of Lemma 2, we consider the specification pattern $u = S*S$, so that $k = 3$ and $s = 2$. The system is then of order 4, and its form is

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

3.2 SINGULAR BEHAVIOR OF THE DIFFERENTIAL SYSTEM. At this stage, we start plunging into complex analysis. Since, by their combinatorial origin, the coefficients $c_{u,n}$ satisfy

$$c_{u,n} = O(n), \tag{11}$$

we thus know that functions $c_u(z)$ and $d_u(z)$ are analytic in the domain $|z| < 1$. From the general theory of linear differential systems[3] in the complex plane, it follows that a solution to system Σ,

$$\frac{d}{dz}[\mathbf{y}(z)] = \Omega(z)\mathbf{y}(z) + \mathbf{b}(z), \tag{12}$$

is analytic in any region where the coefficient matrix Ω and the vector $\mathbf{b}$ are analytic. The only singularities of Ω and $\mathbf{b}$ are at $z = 0$ and $z = 1$. However, we have seen that the solution defined by our initial conditions is analytic around 0. Thus there only remains $z = 1$ as the unique singularity of the vector $\mathbf{y}(z)$. We propose to estimate the coefficients of $d_u(z) \equiv y_1(z)$ by means of the complex integral

$$d_{u,n} = \frac{1}{2i\pi}\int_\Gamma d_u(z)\,\frac{dz}{z^{n+1}}, \tag{13}$$

where Γ is any contour that simply encircles the origin inside the domain of analycity of $d_u(z)$. Following [18] and [7], we propose to choose for Γ a contour that comes close to the singularity $z = 1$. To evaluate the integral (13) then requires detailed expansions for the solutions to system Σ around this point. The matrix $\Omega(z)$ being meromorphic with a single pole at $z = 1$, the homogeneous system (defined by setting $\mathbf{b}$ to 0 in (12)) has what is known as a *singularity of the first kind* and the y_j's are expected to have a logarithmic singularity at $z = 1$. We shall see that the dominant contribution in the local expansion of $d_u(z)$ is of the form

$$d_u(z) \sim \delta \cdot (1 - z)^\lambda \qquad \text{as} \quad z \to 1, \tag{14}$$

with λ the smallest root of the *indicial equation*

$$\det(\Omega_0 - \lambda I) = 0, \tag{15}$$

where I is the $(2k - s) \times (2k - s)$ identity matrix and

$$\Omega_0 = \lim_{z \to 1}(z - 1)\Omega(z).$$

[3] Here and in what follows, we refer to the book by Henrici [11, chap. 9] as our main source on differential systems.

The use of an appropriate contour Γ in (13) shows that we can translate the approximation of a function (14) into an approximation for its Taylor coefficients[4]

$$d_{u,n} \sim [z^n]\delta(1 - z)^{\lambda}. \tag{16}$$

Now, the asymptotics of the coefficients of the right-hand side of (16) is well known; thus, provided that δ is nonzero,

$$d_{u,n} \sim \frac{\delta}{\Gamma(-\lambda)} n^{-\lambda-1}, \tag{17}$$

with $\Gamma(s)$ denoting the Euler gamma function. Theorem 1 then follows from the explicit form of the indicial equation (15).

To proceed with this program, we now prove the key proposition that describes the behavior of function $d_u(z)$.

PROPOSITION 1. *Around $z = 1$, the function $d_u(z)$ has an expansion of the form*

$$d_u(z) = \frac{h_{\alpha_1}}{(1 - z)^{\alpha_1}} + \sum_{\alpha \in I \setminus \{\alpha_1\}} \frac{g_\alpha(\log(z - 1))}{(1 - z)^{\alpha}} + O\left(\frac{1}{(1 - z)^2}\right),$$

where I is the set of all complex roots α of the equation

$$\alpha^s(\alpha - 1)^{k-s} - 2^k = 0$$

satisfying $Re(\alpha) \geq 2$,

$$\alpha_1 = max(I),$$

and each $g_\alpha(u)$ is a polynomial of degree at most 5.

We summarize here the discussion of the proof; details can be filled in by referring to the extensive treatment given by Henrici [11]. The general solution of the *nonhomogeneous system* Σ is the sum of a *particular solution* and of the general solution of the *homogeneous system*

$$\frac{d}{dz}[\mathbf{w}(z)] = \Omega(z)\mathbf{w}(z). \tag{18}$$

We thus study separately the solutions to the homogeneous system (Lemma 3) and then construct a particular solution (Lemma 4).

There is, however, a difficulty that arises in this process: In differential systems, logarithmic terms may be introduced when some confluences occur in expansions. As we shall see, the distinction is based on the roots of the indicial equation (15) and complications occur when two such roots differ by an integer. We need to distinguish two cases (labeled A and B) in Lemmas 3 and 4, depending on condition $\mathscr{H}_{k,s}$:

$$\mathscr{H}_{k,s}\colon \forall \lambda \neq \lambda'[\chi(\lambda) = 0 \text{ and } \chi(\lambda') = 0 \Rightarrow \lambda - \lambda' \notin \mathbb{Z}],$$

where the polynomial $\chi(\lambda)$ related to the indicial equation (15) is defined by

$$\chi(\lambda) = (-\lambda)^s(-1 - \lambda)^{k-s} - 2^k.$$

This condition is satisfied for instance by all integers k, s: $0 < s < k \leq 10$.

[4] We let $[z^n]f(z)$ denote, as usual, the coefficient of z^n in the Taylor expansion of $f(z)$.

LEMMA 3A. *If k and s satisfy condition $\mathscr{H}_{k,s}$, then, around $z = 1$, any solution of the homogeneous system* (18) *has an expansion of the form*

$$\mathbf{w}(z) = \sum_{\alpha \in I} \frac{\mathbf{h}_\alpha}{(1-z)^\alpha} + O\left(\frac{1}{(1-z)^2}\right)$$

for some constant vectors $\mathbf{h}_\alpha$.

PROOF. A fundamental matrix W of system (18) is defined as a matrix whose columns form a linearly independent set of solutions, and thus it satisfies the matrix differential system

$$\frac{d}{dz} W(z) = \Omega(z) W(z). \tag{19}$$

The matrix $\Omega(z)$ is meromorphic at $z = 1$ and we can write

$$\Omega(z) = \frac{1}{z-1} \sum_{m \geq 0} \Omega_m (z-1)^m.$$

The matrix Ω_0 is in the case of system Σ of the form

$$\Omega_0 = -\left(\begin{array}{c|c} A_0 & C_0 \\ \hline 0 & 0 \end{array}\right)$$

with A_0 a matrix of dimension $k \times k$ whose elements are found from matrix A:

$$A_{0,ii} = 0 \quad \text{if} \quad u_i = S,$$

$$A_{0,ii} = 1 \quad \text{if} \quad u_i = *,$$

$$A_{0,i,i+1 \bmod k} = 2;$$

other elements are all equal to 0. The matrix C_0 is equal to B^T (B defined in Lemma 2).

Returning to our previous example where $u = S*S$, we have for instance

$$\Omega_0 = \begin{pmatrix} 0 & -2 & 0 & 0 \\ 0 & -1 & -2 & +1 \\ -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The characteristic polynomial of matrix Ω_0 is determined by successive expansions along the last $k - s$ rows:

$$\operatorname{char}(\Omega_0) = (-\lambda)^{k-s} \operatorname{char}(-A_0),$$

and a direct calculation from the definition of A_0 shows that

$$\operatorname{char}(-A_0) = (-\lambda)^s (-1-\lambda)^{k-s} - 2^k. \tag{20}$$

This is the polynomial $\chi(\lambda)$ introduced in the definition of $\mathscr{H}_{k,s}$. Since

$$\chi'(\lambda) = (-1)^k \lambda^{s-1} (1+\lambda)^{k-s-1} (k\lambda + s),$$

we directly check that $\chi(\lambda)$ has only simple roots for $s \neq 0, k$. Thus A_0 can be diagonalized, and its block structure shows that the same property holds true for Ω_0. Therefore, for some transition matrix T, we have

$$\Omega_0 = T^{-1} \Delta T,$$

where Δ is diagonal with last $k - s$ diagonal elements equal to zero. The system,

$$W'(z) = \frac{\Omega_0}{z-1} W, \tag{21}$$

can be viewed as an "approximation" to system (19); it has a fundamental matrix of the form

$$W(z) = (z-1)^{\Delta} T^{-1}. \tag{22}$$

Considering the system (19) as a perturbation of system (21), one proves by the method of indeterminate coefficients that (19) has a solution of the form

$$W(z) = P(z)(z-1)^{\Delta}, \tag{23}$$

where P is analytic at $z = 1$ and $P(1) = T^{-1}$ under the condition that no two roots of $\chi(\lambda)$ differ by an integer.

We have assumed here that k and s satisfy this condition $\mathscr{H}_{k,s}$. Expressed differently eq. (23) then means that every component W_j of a solution $\mathbf{w}$ of the homogeneous system (18) has a finite expansion of the form

$$W_j(z) = \sum_{\alpha} \frac{h_{\alpha}^{(j)}(z)}{(1-z)^{\alpha}} + h_0^{(j)}(z) \tag{24}$$

for some functions $h_{\alpha}^{(j)}(z)$ analytic at $z = 1$ where the sum is over α's solution of the equation

$$\chi(-\alpha) = 0 \qquad \text{or} \qquad \alpha^s(\alpha-1)^{k-s} - 2^k = 0 \tag{25}$$

(with again $\chi(\lambda)$ defined by (20)). The term $h_0(z)$ corresponds to the eigenvalue 0 of matrix Ω_0.

To conclude with the proof of Lemma 3A, we therefore only need to study the localization of the exponents α in eq. (25). Since these are zeros of the polynomial

$$\chi(-\alpha) = \alpha^s(\alpha-1)^{k-s} - 2^k,$$

they have to satisfy

$$|\alpha| < 3$$

for all values of k and s, and there is always a unique zero α_1 of $\chi(-\alpha)$ in the interval (2, 3). Furthermore, it is easy to check that all other roots of $\chi(-\alpha)$ have a real part strictly less than α_1. (Actually it can be also proved that, when k is large enough, $\chi(-\alpha)$ has several complex roots whose real parts are in the interval $(2, \alpha_1)$.) We thus obtain the statement of the lemma by selecting in (24) only those terms whose α satisfies $\mathrm{Re}(\alpha) > 2$ and retaining only the first terms $h_{\alpha}^{(j)}(1)$ of the $h_{\alpha}^{(j)}(z)$. □

LEMMA 3B. *If k and s do not satisfy condition $\mathscr{H}_{k,s}$, then around $z = 1$ any solution to the homogeneous system* (18) *has an expansion of the form*

$$\mathbf{w}(z) = \frac{\mathbf{h}_{\alpha_1}}{(1-z)^{\alpha_1}} + \sum_{\alpha \in \Lambda \setminus \{\alpha_1\}} \frac{\mathbf{g}_{\alpha}(\log(z-1))}{(1-z)^{\alpha}} + O\left(\frac{1}{(1-z)^2}\right),$$

where each component $g_{\alpha}^{(j)}(u)$ of $g_{\alpha}(u)$ is a polynomial of degree at most 5 in u.

PROOF. The reduction method [11, theorem 9.5.d, p. 122] transforms a system

$$\frac{d}{dz}\mathbf{w} = \frac{1}{(z-1)} \Omega(z)\mathbf{w},$$

where matrix $\Omega(1)$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ into a system

$$\frac{d}{dz}\,\hat{\mathbf{w}} = \frac{1}{(z-1)}\,\hat{\Omega}(z)\hat{\mathbf{w}}, \tag{26}$$

where matrix $\hat{\Omega}(1)$ has eigenvalues $\lambda_1 - 1, \lambda_2, \ldots, \lambda_m$; the relation between $\mathbf{w}$ and $\hat{\mathbf{w}}$ being of the form

$$W(z) = H(z)\hat{W}(z) \tag{27}$$

for some analytic matrix $H(z)$.

Using it repeatedly, we transform the original system into a system of the form (26) with the correspondence given by (27), in such a manner that (a) the eigenvalues of $\hat{\Omega}(1)$ are a subset of the eigenvalues of $\Omega(1)$; (b) no two eigenvalues of $\hat{\Omega}(1)$ differ by an integer. Furthermore, the "dominant" eigenvalue α_1 still has multiplicity 1; each of the other nonzero eigenvalues has multiplicity at most 6, since any root of $\chi(\lambda)$ has to satisfy $|\lambda| < 3$; finally eigenvalue 0 admits a set of $k - s$ linearly independent eigenvectors.

A fundamental matrix of system (26) can thus be put under the form

$$\hat{W}(z) = \hat{P}(z)(z-1)^{\hat{S}}, \tag{28}$$

with $\hat{P}(z)$ analytic at 1 and $\hat{S}$ upper triangular.

However, we are no longer guaranteed that $\hat{S}$ may be diagonalized. Matrix $\hat{S}$ decomposes into

$$\hat{S} = \hat{\Delta} + \hat{U}$$

with $\hat{\Delta}$ diagonal, $\hat{U}$ a strict upper triangular matrix (i.e., with all its diagonal elements being 0), which commutes with $\hat{\Delta}$ (see [11, p. 120]), and $\hat{U}^6 = 0$ (the null matrix). Thus

$$(z-1)^{\hat{S}} = e^{\hat{S}\log(z-1)} = (z-1)^{\hat{\Delta}}\left(I + \sum_{k=1}^{5} \frac{\hat{U}^k}{k!}\,(\log(z-1))^k\right). \tag{29}$$

Grouping (27)–(29) establishes the claim of the Lemma. □

The next stage now consists in constructing a particular solution of the nonhomogeneous system Σ. This is achieved by means of the matrix "variation-of-constants" formula.

LEMMA 4A. *If hypothesis $\mathscr{H}_{k,s}$ is satisfied, the nonhomogeneous system Σ admits in a neighborhood of $z = 1$ a particular solution of the form*

$$\frac{\mathbf{H}(z)}{(1-z)^2} + \mathbf{G}(z)\,\log(z-1),$$

where $\mathbf{H}(z)$ and $\mathbf{G}(z)$ are analytic at $z = 1$.

PROOF. By the variation-of-constants formula [11, p. 99], if $\hat{z}$ is a regular point of the system (for instance $\hat{z} = \frac{1}{2}$) and $W(z)$ a fundamental matrix of the homogeneous system, the general solution to the nonhomogeneous system is given by

$$\mathbf{w}(z) = W(z)W^{-1}(\hat{z})\mathbf{c} + W(z)\int_{\hat{z}}^{z} W^{-1}(t)\mathbf{b}(t)\,dt, \tag{30}$$

and the second term is a particular solution of the nonhomogeneous system. We know that the homogeneous system has a fundamental matrix of the form

$$W(z) = P(z)(z-1)^{\Delta},$$

where $P(z)$ is analytic at 1:

$$P(z) = \sum_{m\geq 0} P_m(z-1)^m,$$

and P_0 is regular. Thus

$$W^{-1}(z) = (z-1)^{-\Delta}Q(z),$$

where

$$Q(z) = \sum_{m\geq 0} Q_m(z-1)^m,$$

and, again, Q_0 regular. Since

$$\mathbf{b}(z) = \frac{\mathbf{b}_0}{(z-1)^3}$$

for some constant vector $\mathbf{b}_0$, we find, for the particular solution

$$\mathbf{U}(z) = W(z)\int_{\hat{z}}^{z} W^{-1}(t)\mathbf{b}(t)\,dt,$$

the expansion

$$\mathbf{U}(z) = P(z)\left[\sum_{n\geq 0}(z-1)^{\Delta}\int_{\hat{z}}^{z}(t-1)^{-\Delta-n-3}\,dt\;Q_n\mathbf{b}_0\right]. \tag{31a}$$

Integration of the matrix shows that (taking for instance $\hat{z} = \frac{1}{2}$)

$$\mathbf{U}(z) = P(z)\left[\sum_{n\geq 0}(z-1)^{\Delta}(I_n(z) - I_n(\hat{z}))Q_n\mathbf{b}_0\right], \tag{31b}$$

where $I_n(z)$ is a diagonal matrix whose elements are

—if $n \neq 2$

$$\frac{(z-1)^{-\lambda_j+n-2}}{-\lambda_j+n-2}, \qquad 1\leq j\leq k,$$

$$\frac{(z-1)^{n-2}}{n-2}, \qquad k<j\leq 2k-s;$$

—if $n = 2$

$$\frac{(z-1)^{-\lambda_j}}{-\lambda_j}, \qquad 1\leq j\leq k,$$

$$\log(z-1), \qquad k<j\leq 2k-s, \tag{32}$$

with $\lambda_1, \lambda_2, \ldots, \lambda_k$ the roots of polynomial $\chi(\lambda)$.

Splitting the sum in (31b), we find

$$\mathbf{U}(z) = \mathbf{U}_1(z) - \mathbf{U}_2(z),$$

where

$$\mathbf{U}_1(z) = P(z)\sum_{n\geq 0}(z-1)^{\Delta}I_n(z)Q_n\mathbf{b}_0,$$

$$\mathbf{U}_2(z) = P(z)\sum_{n\geq 0}(z-1)^{\Delta}I_n(\hat{z})Q_n\mathbf{b}_0.$$

The vector $\mathbf{U}_2(z)$ is a solution of the homogeneous system, so that a particular solution of system Σ is provided by $\mathbf{U}_1(z)$. Separating the terms in the sum according to $n \neq 2$, $n = 2$, we have

$$\mathbf{U}_1(z) = P(z) \sum_{n \neq 2} (z-1)^{\Delta} I_n(z) Q_n \mathbf{b}_0 + P(z)(z-1)^{\Delta} I_2(z) Q_2 \mathbf{b}_0. \tag{33}$$

The diagonal form of $I_n(z)$ in (32) shows that terms of the form $(z-1)^{\pm\lambda}$ disappear in the products of (33) and we are left with

$$\mathbf{U}_1(z) = \frac{\mathbf{H}(z)}{(z-1)^2} + \mathbf{G}(z)\log(z-1)$$

for some vectors $\mathbf{H}$ and $\mathbf{G}$ analytic at $z = 1$. □

LEMMA 4B. *If hypothesis $\mathscr{H}_{k,s}$ is not satisfied, then the nonhomogeneous system Σ admits in a neighborhood of $z = 1$ a particular solution of the form*

$$\frac{\mathbf{H}(z)}{(1-z)^2} + \sum_{k=1}^{5} \mathbf{G}_k(z)(\log(z-1))^k,$$

where $\mathbf{H}(z)$ and the $\mathbf{G}_k(z)$ are analytic at $z = 1$.

PROOF. The previous method applied to this case would rather trivially imply the existence of a particular solution with dominant terms of the form

$$\frac{(\log(z-1))^5}{(z-1)^2}.$$

However, the stronger property of the statement of the lemma is required for the later part of the analysis. It is derived by what looks like a "failed attempt" at a direct solution of system Σ by the method of indeterminate coefficients.

Let $\mathbf{H}(z)$ have the expansion

$$\mathbf{H}(z) = \sum_{m \geq 0} \mathbf{H}_m (z-1)^m.$$

If we try to identify coefficients of $\mathbf{H}(z)$ so that

$$\frac{\mathbf{H}(z)}{(z-1)^2}$$

satisfies system Σ, we find the equations

$$(\Omega_0 + 2I)\mathbf{H}_0 = -\mathbf{b}_0, \qquad (\Omega_0 + I)\mathbf{H}_1 = -\Omega_1 \mathbf{H}_0, \tag{34}$$

where $\mathbf{b}_0$ is a constant vector defined by

$$\mathbf{b}(z) = \frac{\mathbf{b}_0}{(z-1)^3}.$$

System (34) is solvable since $(\Omega_0 + 2I)$ and $(\Omega_0 + I)$ are nonsingular. The next equation would be

$$\Omega_0 \mathbf{H}_2 = -\Omega_1 \mathbf{H}_1 - \Omega_2 \mathbf{H}_0,$$

which need not be solvable since Ω_0 is singular. However if $\mathbf{w}(z)$ is a solution to system Σ and $\mathbf{H}_0$, $\mathbf{H}_1$ are defined by (34), we find that

$$\overline{\mathbf{w}}(z) = \mathbf{w}(z) - \frac{\mathbf{H}_0}{(z-1)^2} - \frac{\mathbf{H}_1}{(z-1)} \tag{35}$$

satisfies the modified system

$$\frac{d}{dz}\overline{\mathbf{w}}(z) = \Omega(z)\overline{\mathbf{w}}(z) + \overline{\mathbf{b}}(z), \tag{36}$$

where $\overline{\mathbf{b}}(z)$ has now only a simple pole at $z = 1$. It is to the transformed system (36) that we now apply the method of variation of constants. By the developments of Lemma 3B, a fundamental matrix of the homogeneous system corresponding to (36) is of the form

$$W(z) = P(z)(z-1)^{\hat{S}} = P(z)(z-1)^{\hat{\Delta}}\left(I + \sum_{k=1}^{5} \frac{\hat{U}^k}{k!}(\log(z-1))^k\right)$$

and its inverse may be similarly written

$$W^{-1}(z) = \left[I + \sum_{k=1}^{5} (-1)^k \frac{\hat{U}^k}{k!}(\log(z-1))^k\right](z-1)^{-\hat{\Delta}}Q(z).$$

We can use, as in Lemma 4A, this form in the variation-of-constants formula. A particular solution to (36) is thus given by

$$\overline{\mathbf{w}}(z) = P(z)(z-1)^{\hat{\Delta}}X \sum_{n\geq 0} \int_{\hat{z}}^{z} Y(t-1)^{-\hat{\Delta}+(n-1)I}\, dt\, Q_n\overline{\mathbf{b}}_0$$

for some vector of constants $\overline{\mathbf{b}}_0$ and some matrices X, Y whose coefficients are polynomial in $\log(z-1)$; X and Y also commute with $\hat{\Delta}$ or with matrices of a similar block structure like $(z-1)^{\pm\Delta}$. Carrying out the integration explicitly leads to

$$\overline{\mathbf{w}}(z) = \overline{\mathbf{H}}(z) + \sum_{k=1}^{5} \mathbf{G}_k(z)\log(z-1)^k,$$

which, combined with (35), yields the claim of the lemma. □

We can now conclude the proof of Proposition 1: The most general solution to system Σ is obtained as a sum of the particular solution $\mathbf{w}(z)$ constructed in Lemma 4 that satisfies

$$\mathbf{w}(z) = O\left(\frac{1}{(z-1)^2}\right)$$

and of the general solution to the homogeneous system whose behavior is described in Lemma 3.

3.3 Asymptotics of Coefficients. The next stage consists in translating the expansion of $d_u(z)$ around its singularity $z = 1$ into information about the asymptotics of its coefficients. This uses the following results:

PROPOSITION 2

(*i*) *The nth Taylor coefficient* c_n *of the function*

$$c(z) = (1 - z)^{-\alpha}[\log(1 - z)]^k$$

satisfies

$$c_n = n^{\alpha-1}\Pi(\log n) + O(n^{\alpha-2}\log^k n)$$

for some polynomial Π (*depending on* α *and* k) *of degree at most* k.

(*ii*) *Suppose that* $g(z)$ *is analytic in*

$$E = \{z: |z| \leq 1, z \neq 1\}$$

and that for $z \in E$

$$g(z) = O(|1 - z|^{-\beta})$$

for some $\beta > 1$. *Then the nth Taylor coefficient* g_n *of* $g(z)$ *satisfies*

$$g_n = O(n^{\beta-1}).$$

PROOF. Part (ii) of the proposition is taken from [7, Prop. 7, p. 209]: It is proved there by expressing c_n by means of the Cauchy integral formula and taking as a contour of integration the circle of convergence of $g(z)$, except for a small notch inside the circle at distance $1/n$ of the singularity 1. (See also [18] for related results.) Proof of part (i) of the proposition relies on similar methods, except that now precise asymptotic results are needed. One starts from the integral form of c_n,

$$c_n = \frac{1}{2i\pi}\int_\Gamma c(z)\frac{dz}{z^{n+1}}, \tag{37}$$

and uses for Γ the contour (oriented counterclockwise)

$$\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4,$$

where

$$\Gamma_1 = \left\{z = 1 - \frac{e^{i\theta}}{n}: \theta \in \left[-\frac{\pi}{2}; +\frac{\pi}{2}\right]\right\},$$

$$\Gamma_2 = \left\{z = 1 + \frac{i}{n} + \frac{x}{n}: x \in [0; n]\right\},$$

$$\Gamma_3 = \left\{z: |z| = \left(4 + \frac{1}{n^2}\right)^{1/2}, \operatorname{Re}(z) < 2\right\},$$

$$\Gamma_4 = \{z: \bar{z} \in \Gamma_2\}.$$

This contour is depicted in Figure 2. Decomposing the integral (37) along the particular contour Γ, we have

$$c_n = c_n^{(1)} + c_n^{(2)} + c_n^{(3)} + c_n^{(4)}.$$

Since $c(z)$ is bounded along Γ_3,

$$c_n^{(3)} = O(2^{-n}), \tag{38}$$

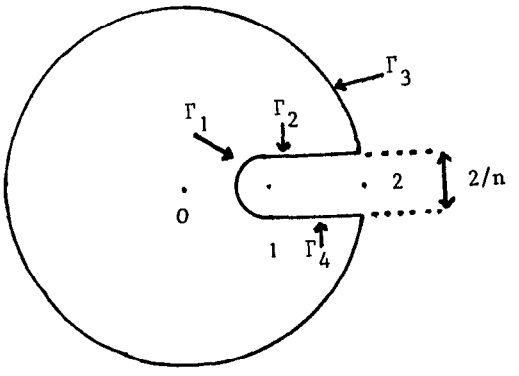


FIG. 2. The contour Γ used in the proof of Proposition 2.

so that we only have to evaluate $c_n^{(1)}$ on the one hand and $c_n^{(2)}$, $c_n^{(4)}$ on the other. As we proceed with the evaluation of $c_n^{(1)}$, the change of variable

$$z = 1 - \frac{e^{i\theta}}{n}$$

shows that

$$c_n^{(1)} = -\frac{n^{\alpha-1}}{2\pi} \int_{-\pi/2}^{+\pi/2} (i\theta - \log n)^k e^{-i(\alpha-1)\theta} \left(1 - \frac{e^{i\theta}}{n}\right)^{-(n+1)} d\theta.$$

Using the exponential approximation

$$\left(1 - \frac{e^{i\theta}}{n}\right)^{-(n+1)} = e^{e^{i\theta}} + O\left(\frac{1}{n}\right)$$

in the previous integral, we find

$$c_n^{(1)} = -\frac{n^{\alpha-1}}{2\pi} \int_{-\pi/2}^{+\pi/2} (i\theta - \log n)^k e^{-i(\alpha-1)\theta} e^{e^{i\theta}} d\theta + O((\log n)^k n^{\alpha-2}). \quad (39)$$

The integral there is clearly a polynomial of degree at most k in $\log n$. For the integral along Γ_2, we use the change of variable

$$z = 1 + \frac{i}{n} + \frac{x}{n},$$

with which we find

$$c_n^{(2)} = \frac{n^{\alpha-1}}{2i\pi} \int_0^n \frac{(\log(-i - x) - \log n)^k}{(-i - x)^\alpha} \left(1 + \frac{x + i}{n}\right)^{-(n+1)} dx.$$

Similarly, using the exponential approximation for $x \le n^{1/3}$,

$$\left(1 + \frac{x + i}{n}\right)^{-(n+1)} = e^{-i-x}\left(1 + O\left(\frac{1 + x^2}{n}\right)\right),$$

in $c_n^{(2)}$ (the integral from $n^{1/3}$ to n is exponentially small), we find

$$c_n^{(2)} = \frac{n^{\alpha-1}}{2i\pi} \int_0^{n^{1/3}} \frac{(\log(-i - x) - \log n)^k}{(-i - x)^\alpha} e^{-i-x} dx + O((\log n)^k n^{\alpha-2}).$$

The integral can be extended from 0 to ∞ introducing only exponentially small terms, so that

$$c_n^{(2)} = n^{\alpha-1} \frac{e^{-i}}{2i\pi} \int_0^\infty \frac{(\log(-i-x) - \log n)^k}{(-i-x)^\alpha} e^{-x}\, dx + O((\log n)^k n^{\alpha-2}), \quad (40)$$

the integral being again a polynomial in $\log n$ of degree at most k. The case of $c_n^{(4)}$ is entirely similar, and combining (38), (39), and (40) establishes the claim of Proposition 2. Notice also that the same method would make it possible to determine a complete asymptotic expansion of c_n for any fixed k and fixed α. Finally, note that the polynomials Π cannot be identically zero. □

Now a direct application of Proposition 2 to the result of Proposition 1 shows that $d_{u,n}$ admits the asymptotic expansion

$$d_{u,n} = \frac{h_{\alpha_1}}{\Gamma(\alpha_1)} n^{\alpha_1-1} + \sum_{\alpha \in I \setminus \{\alpha_1\}} \xi_\alpha(\log n) n^{\alpha-1} + O(n),$$

where the $\xi_\alpha(u)$ are polynomials of degree at most 5. Since

$$c_{u,n} = \frac{1}{n+1} d_{u,n},$$

we thus have

$$c_{u,n} = \frac{h_{\alpha_1}}{\Gamma(\alpha_1)} n^{\alpha_1-2} + \sum_{\alpha \in I \setminus \{\alpha_1\}} \xi_\alpha(\log n) n^{\alpha-2} + O(1). \quad (41)$$

To complete the proof of Theorem 1, we therefore only need to show that the actual order of $c_{u,n}$ is given by the first term of (41):

$$c_{u,n} = \Omega(n^{\alpha_1-2}),$$

or, equivalently,

$$h_{\alpha_1} \neq 0.$$

LEMMA 5. *The coefficient h_{α_1} in the expansion of $c_{u,n}$ is strictly positive.*

PROOF. The proof, which is nonconstructive, proceeds through an indirect argument using the positivity of the $c_{u,n}$ and a logarithmic lower bound on the $c_{u,n}$. Assume *a contrario* that $h_{\alpha_1} = 0$.

(i) If all the h_α were equal to zero, when we would have $c_{u,n} = O(1)$, $d_{u,n} = O(n)$ as $n \to \infty$.

This contradicts the fact that $c_{u,n}$ is at least as large as the cost of a completely specified search which is known to be $O(\log n)$. The analytic equivalent of this argument consists in observing that

$$d_{u,n} \geq d_{\sigma n} \qquad \text{where} \quad \sigma = S\,S \cdots S, \quad |\sigma| = k,$$

as follows from recurrences (2) and (4). Then the solution of the equation for $d_\sigma(z)$,

$$d'_\sigma(z) = \frac{2}{(1-z)^3} + \frac{2}{1-z} d_\sigma(z),$$

is found to be

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

so that

$$d_{\sigma,n} \sim 2n \log n,$$

a contradiction in this case.

(ii) Thus, if $h_{\alpha_1} = 0$, at least one of the h_α for $\alpha \in I \setminus \{\alpha_1\}$ is nonzero. The complex roots of the indicial equation

$$\chi(-\alpha) = 0$$

occur in pairs of complex conjugates. Let therefore β, $\bar{\beta}$ be the roots of highest real part such that

$$h_\beta \neq 0, \qquad h_{\bar{\beta}} = \bar{h}_\beta \neq 0;$$

from Proposition 2 it follows that for some constant $C \neq 0$

$$c_{u,n} \sim Cn^{\beta-2}(\log n)^k + \bar{C}n^{\bar{\beta}-2}(\log n)^k,$$

for some k where $0 \leq k \leq 5$. Thus with

$$C = a + ib, \qquad \beta = \sigma + it,$$

we find

$$c_{u,n} \sim 2n^{\sigma-2}(\log n)^k(a \cos(t \log n) - b \sin(t \log n)).$$

But such an equation contradicts the fact that the $c_{u,n}$ are nonnegative numbers.

We have thus seen that in all cases the assumption $h_{\alpha_1} = 0$ leads to a contradiction, so that Lemma 5 is established. □

Lemma 5 itself allows us to complete the proof of Theorem 1; the dominant exponent e in the asymptotic form of $c_{u,n}$,

$$c_{u,n} \sim \gamma_u n^e,$$

is $e = \alpha_1 - 2$, and it is the unique positive real root of the equation

$$\chi(-2 - e) = 0;$$

that is,

$$(e + 2)^s(e + 1)^{k-s} - 2^k = 0;$$

or, equivalently,

$$(e + 2)^{s/k}(e + 1)^{1-s/k} - 2 = 0,$$

and the function $\theta(s/k)$ is

$$\theta\left(\frac{s}{k}\right) = e - \left(1 - \frac{s}{k}\right),$$

which therefore satisfies the equation of the statement of Theorem 1.

4. *Digital Techniques for Internal and External Search*

In this section, we provide an analysis of partial match retrieval for *k-d-tries* (in Section 4.1) and for *grid file* algorithms (Section 4.2). Our basic interest is in the so-called *Bernoulli model* corresponding to the description given in Section 2: The number of keys in the file is a fixed integer n and keys are assumed to be taken independently from a uniform distribution. As a consequence of these hypotheses, bits of arbitrary positions in arbitrary fields of keys are independent uniform

{0, 1} random variables. There is also strong interest in a closely related model, called the *Poisson model* (see, for instance, [5] for analyses under this model); there, the number of keys in the file is assumed to be a random variable N with a Poisson distribution, such that

$$\Pr(N = k) = e^{-n}\frac{n^k}{k!}$$

for some fixed parameter n that corresponds to the expectation of N. The interest in the Poisson model is that it can make certain technical developments simpler because of certain strong independence properties of the localization of keys in nonoverlapping subintervals.

We have analyzed k-d-tries and grid files under both the Bernoulli and Poisson models, and the results are asymptotically equivalent. For the sake of conciseness, we illustrate the analytic techniques involved by giving only the proof of the evaluation of k-d-tries under the Bernoulli model and of grid-file algorithms under the Poisson model.

4.1 MULTIDIMENSIONAL TRIES. Again consider a file $F \subset D_1 \times D_2 \times \cdots \times D_k$, where each attribute domain D_i is assimilated to the set of infinite binary sequences

$$D_i \cong \{0, 1\}^\infty.$$

With any record $r = (r_1, r_2, \ldots, r_k)$ there is associated an infinite binary sequence in the usual manner through *regular shuffling*. Let

$$r_j = r_j^{(1)}, r_j^{(2)}, r_j^{(3)}, \ldots \quad \text{where} \quad r_j^{(k)} \in \{0, 1\}$$

be the binary representation of attribute r_j; the infinite sequence associated with r is

$$\rho = \text{shuffle}(r) = \rho^{(1)}, \rho^{(2)}, \rho^{(3)}, \ldots \quad \text{where} \quad \rho^{(k)} \in \{0, 1\},$$

where

$$\rho = r_1^{(1)}, r_2^{(1)}, \ldots, r_k^{(1)}, r_1^{(2)}, r_2^{(2)}, \ldots, r_k^{(2)}, r_3^{(1)}, r_2^{(3)}, \ldots, r_k^{(3)}, \ldots.$$

Thus, the shuffle of a k-tuple is obtained by taking in sequence the first bit of attribute 1, the first bit of attribute 2, ..., the first bit of attribute k, and then starting cyclically again with the second bits of attributes 1, 2, ..., k, etc.

By definition, the *k-d-trie* constructed on a finite set F is the (1-d-) trie constructed on the set $\{\text{shuffle}(r)/r \in F\}$. Thus, k-d-tries have some analogy to k-d-trees with the notable difference that the partitioning of elements corresponds to fixed values of the fields instead of to values provided by the file itself, and records are stored at the leaves of the tree. The fact that 1-d-tries tend to be better balanced than 1-d-search trees does not crucially affect the performances of one-dimensional search, which are logarithmic in both cases. However, in the context of multidimensional search, it leads to asymptotically smaller orders, as we now prove.

THEOREM 2. *The average cost, measured by the number of internal nodes traversed, of a partial match query of specification pattern u with s specified attributes in a k-d-trie constructed from a file of either size n (under the Bernoulli model) or expected size n (under the Poisson model) satisfies*

$$c_{u,n} = \gamma\left(\frac{1}{k}\log_2 n\right)n^{1-s/k} + O(1),$$

where $\gamma(u)$ is a periodic function of u with period 1, small amplitude, and mean value

$$\gamma_0 = -\frac{s}{k^2 \log 2}\,\Gamma\left(\frac{s}{k} - 1\right) \sum_{l=0}^{k-1} (\delta_1\, \delta_2 \cdots \delta_l) 2^{-l(1\,-s/k)}$$

with $\delta_l = 1$ if the lth attribute of the query is specified, and $\delta_l = 2$ if it is unspecified.

As announced earlier, we only give here the proof of the estimate under the Bernoulli model. The proof under the Poisson model follows trivially by adapting the methods introduced in Section 4.2.

LEMMA 6. *The exponential generating function of the average costs $c_{u,n}$ under the Bernoulli model,*

$$c_u(z) = \sum_{n \geq 0} c_{u,n} \frac{z^n}{n!},$$

satisfies the relation

$$c_u(z) = \delta_1 e^{z/2} c_{u'}\left(\frac{z}{2}\right) + e^z - 1 - z,$$

with u' obtained by circularly shifting the letters of u by one position to the left.

PROOF. Let

$$t = t_1 \hat{} t_2$$

be a k-d-trie associated to a particular file F. If the first attribute of the query is nonspecified, we have, for the expected cost of a random query,

$$c_u[t] = 1 + c_{u'}[t_1] + c_{u'}[t_2], \tag{42}$$

since the search then has to proceed in parallel along both subtrees with attributes changing cyclically according to pattern u'. If, contrariwise, the first attribute is specified, we find

$$c_u[t] = 1 + \tfrac{1}{2}(c_{u'}[t_1] + c_{u'}[t_2]), \tag{43}$$

since with probability $\frac{1}{2}$ the first bit of the first attribute of the query starts with a 0 (the search then proceeds in t_1) and with probability $\frac{1}{2}$ it starts with a 1 (the search then proceeds in t_2).

Given n random elements $n \geq 2$ organized in a k-d-trie $t = t_1 \hat{} t_2$, the probability that

$$|t_1| = p, \qquad |t_2| = n - p$$

is given by the Bernoulli probabilities

$$\binom{n}{p}\left(\frac{1}{2}\right)^p \left(\frac{1}{2}\right)^{n-p} = \frac{1}{2^n}\binom{n}{p},$$

whence, for the expected values, the recurrences

$$\text{if} \quad u = *v, \qquad c_{u,n} = 1 + \frac{2}{2^n} \sum_p \binom{n}{p} c_{u',p}, \qquad n \geq 2;$$

$$\text{if} \quad u = Sv, \qquad c_{u,n} = 1 + \frac{1}{2^n} \sum_p \binom{n}{p} c_{u',p}, \qquad n \geq 2.$$

In general, for all u and n, we therefore have

$$c_{u,n} = 1 + \frac{\delta_1}{2^n} \sum_p \binom{n}{p} c_{u',p} - \delta_{n,0} - \delta_{n,1}. \tag{44}$$

The translation of (44) in terms of exponential generating functions yields the claim of the lemma. □

LEMMA 7. *The generating function $c_u(z)$ satisfies the difference equation*

$$c_u(z) = 2^{k-s} exp\left[z\left(1 - \frac{1}{2^k}\right)\right] c_u\left(\frac{z}{2^k}\right) + \sum_{j=0}^{k-1} (\delta_1\, \delta_2 \cdots \delta_j) exp\left[z\left(1 - \frac{1}{2^j}\right)\right] \cdot \left(exp\left(\frac{z}{2^j}\right) - 1 - \frac{z}{2^j}\right).$$

PROOF. From Lemma 6, we see that $c_u(z)$ is the first component of a vectorial system of difference equations:

$$\text{System } \Sigma: \begin{cases} c_u(z) = \delta_1 e^{z/2} c_{u'}\left(\frac{z}{2}\right) + e^z - 1 - z, \\ c_{u'}(z) = \delta_2 e^{z/2} c_{u''}\left(\frac{z}{2}\right) + e^z - 1 - z, \\ \vdots \\ c_{u^{(k-1)}}(z) = \delta_k e^{z/2} c_u\left(\frac{z}{2}\right) + e^z - 1 - z. \end{cases}$$

This system can be solved by successive eliminations. Let $a(z)$ denote $e^z - 1 - z$. Transporting the expression of $c_{u'}(z)$ given by the second equation inside the defining equation for $c_u(z)$, we find

$$c_u(z) = a(z) + \delta_1 e^{z/2} a\left(\frac{z}{2}\right) + \delta_1 \delta_2 e^{z/2} e^{z/4} c_{u''}\left(\frac{z}{4}\right).$$

We continue in this fashion, using the equation satisfied by $c_{u''}$, $c_{u'''}$ until the relation is only in terms of $c_u(z)$ itself. □

A functional equation of the form satisfied by $c_u(z)$, namely,

$$\phi(z) = \alpha e^{\beta z} \phi(\gamma z) + A(z)$$

(with ϕ the unknown function), may be solved formally by iteration in a manner similar to the proof of Lemma 7

$$\begin{aligned} \phi(z) &= A(z) + \alpha e^{\beta z} A(\gamma z) + \alpha^2 e^{(\beta+\beta\gamma)z} \phi(\gamma^2 z) \\ &= A(z) + \alpha e^{\beta z} A(\gamma z) + \alpha^2 e^{(\beta+\beta\gamma)z} A(\gamma^2 z) \\ &\quad + \alpha^3 \exp((\beta + \beta\gamma + \beta\gamma^2)z) \phi(\gamma^3 z) \\ &\quad \vdots \\ &= \sum_{j \geq 0}^{\infty} \alpha^j exp\left(\beta\left(\frac{1-\gamma^j}{1-\gamma}\right)z\right) A(\gamma^j z). \end{aligned}$$

Thus using here the particular form of α, β, γ, and $A(z)$ we find

$$c_u(z) = \sum_{j=0}^{\infty} 2^{j(k-s)}\left\{\left[\exp(z) - \exp\left(z\left(1 - \frac{1}{2^{kj}}\right)\right)\left(1 + \frac{z}{2^{kj}}\right)\right]\right.$$
$$+ \delta_1\left[\exp(z) - \exp\left(z\left(1 - \frac{1}{2.2^{kj}}\right)\right)\left(1 + \frac{z}{2.2^{kj}}\right)\right]$$
$$\left. + \delta_1\delta_2\left[\exp(z) - \exp\left(z\left(1 - \frac{1}{4.2^{kj}}\right)\right)\left(1 + \frac{z}{4.2^{kj}}\right)\right] + \cdots\right\},$$

where inside the infinite summation we have a sum of k terms.

Extracting the Taylor coefficients of $c_u(z)$ given by this sum, we get

LEMMA 8. *The expected cost of a partial match retrieval has for $n \geq 2$ the explicit form*

$$c_{u,n} = \sum_{l=0}^{k-1} \delta_1\ \delta_2 \cdots \delta_l \sum_{j\geq 0} 2^{j(k-s)}\tau_{j,l}(n), \tag{45}$$

where for j and l not both zero

$$\tau_{j,l}(x) = 1 - (1 - 2^{-kj-l})^x - x2^{-kj-l}(1 - 2^{-kj-l})^{x-1} \tag{46}$$

and

$$\tau_{0,0}(x) = 1.$$

We observe that the convergence of (45) is guaranteed by the fact that, for fixed n, as j tends to infinity

$$\tau_{j,l}(n) \sim 1 - \exp(-n2^{-kj-l}) - n2^{-kj-l}\exp(-n2^{-kj-l}) = O(n^2 2^{-2kj}). \tag{47}$$

Indeed the exponential approximation (47) is usually the starting point of asymptotic evaluations, but here we shall use a different approach (see [19] for other applications), which is more direct and may be used to obtain asymptotic expansions to any order if required.

We also observe that each $\tau_{j,l}$ is a positive number at most 1, so that if we sum on $j = 1$ to ∞ in (45), we introduce an error term that is bounded above by $k.2^{s-k}$:

$$c_{u,n} = \phi(n) + O(1),$$

where

$$\phi(x) = \sum_{l=0}^{k-1} \delta_1\ \delta_2 \cdots \delta_l \sum_{j\geq 1} 2^{j(k-s)}\tau_{j,l}(x). \tag{48}$$

Equations (46) and (48) thus define $\phi(x)$ for arbitrary real $x \geq 0$. We propose to perform the asymptotic analysis of $\phi(x)$ by investigating properties of its Mellin transform given by

$$\phi^*(\sigma) = \int_0^{\infty} \phi(x)x^{\sigma-1}\,dx. \tag{49}$$

It is known (see, for instance, [3] and [4]) that under suitable analytic conditions, the asymptotic properties of $\phi(x)$ as $x \to \infty$ are directly related to the singularities of $\phi^*(\sigma)$ in a right half plane. We therefore need to derive an expression for $\phi^*(\sigma)$

that reveals some of its singularities and provides an analytic continuation of the integral definition (49). We prove

PROPOSITION 3. *The Mellin transform of the function* $\phi(x)$ *given by equation* (48), *such that*

$$c_{u,n} = \phi(n) + O(1),$$

has the form

$$\phi^*(\sigma) = -(1 + \sigma)\Gamma(\sigma)\left[\frac{2^{k(\sigma-\sigma_0)}}{1 - 2^{k(\sigma-\sigma_0)}} + A(\sigma)\right] \sum_{l=0}^{k-1} \delta_1 \delta_2 \cdots \delta_l 2^{l\sigma},$$

where $\sigma_0 = -(1 + s/k)$ *and* $A(\sigma)$ *is analytic in* $-1 \leq Re(\sigma) \leq s/(2k)$ *and satisfies in this region*

$$A(\sigma) = O(|\sigma|^2), \qquad |\sigma| \to \infty.$$

PROOF. We appeal to the following classical properties of Mellin transforms:

(i) $$\int_0^\infty (e^{-x} - 1)x^{\sigma-1}\, dx = \Gamma(\sigma), \qquad -1 < \mathrm{Re}(s) < 0;$$

(ii) $$\int_0^\infty (xe^{-x})x^{\sigma-1}\, dx = \sigma\Gamma(\sigma), \qquad -1 < \mathrm{Re}(s);$$

(iii) $$\int_0^\infty f(ax)x^{\sigma-1}\, dx = a^{-\sigma} \int_0^\infty f(x)x^{\sigma-1}\, dx, \qquad a > 0.$$

Writing $\tau_{jl}(x)$ under the form

$$\tau_{jl}(x) = 1 - \exp(-x\alpha_{jl}) - \beta_{jl}x \exp(-x\alpha_{jl}),$$

with

$$\alpha_{jl} = -\log(1 - 2^{-kj-l}),$$
$$\beta_{jl} = 2^{-kj-l}(1 - 2^{-kj-l})^{-1},$$

we find thus that the Mellin transform of $\tau_{jl}(x)$ is

$$\tau_{jl}^*(\sigma) = -(\alpha_{jl})^{-\sigma}\Gamma(\sigma) - \beta_{jl}(\alpha_{jl})^{-\sigma-1}\sigma\Gamma(\sigma), \tag{50}$$

provided $-1 < \mathrm{Re}(\sigma) < 0$. From (50), we can determine the expression of $\phi^*(\sigma)$ applying the linearity of the transform to the defining equation (48). The conditions on the values of σ, in order for the interchange of integration in (49) and the infinite summation in (48) to be justified, are that the sums

$$\omega_l(\sigma) = \sum_{j\geq 1} 2^{j(k-s)}(\alpha_{jl})^{-\sigma}, \qquad \omega_l'(\sigma) = \sum_{j\geq 1} 2^{j(k-s)}\beta_{jl}(\alpha_{jl})^{-\sigma-1} \tag{51}$$

be absolutely convergent. Using the asymptotic equivalents

$$(\alpha_{jl})^{-\sigma} = O(2^{kj\,\mathrm{Re}(\sigma)}), \qquad \beta_{jl} = O(2^{-kj}),$$

we see that the sums defining $\omega_l(\sigma)$ and $\omega_l'(\sigma)$ are uniformly and absolutely convergent when σ is in any stripe:

$$S_\eta\colon -1 < \mathrm{Re}(\sigma) < -\left(1 - \frac{s}{k}\right) - \eta, \qquad \eta > 0. \tag{52}$$

Thus the transform of $\phi(x)$ is defined in the S_0 stripe and there

$$\phi^*(\sigma) = -\sum_{l=0}^{k-1} \delta_1\, \delta_2 \cdots \delta_l(\omega_l(\sigma) + \sigma\omega_l'(\sigma)) \cdot \Gamma(\sigma). \tag{53}$$

The next stage consists in analytically continuing $\phi^*(\sigma)$, that is to say, the $\omega_l(\sigma)$, to a domain that extends to the right of $\sigma_0 = -(1 - s/k)$. To that purpose we use the expansion valid for small u uniformly in σ for σ in any fixed stripe $c < \mathrm{Re}(\sigma) < d$:

$$(-\log(1-u))^{-\sigma} = u^{-\sigma}\left(1 - \frac{\sigma u}{2} + O(|\sigma|^2 u^2)\right). \tag{54}$$

This expansion suggests "approximating" $\omega_l(\sigma)$ and $\omega_l'(\sigma)$ by the series

$$\hat{\omega}_l(\sigma) = \sum_{j\geq 1} 2^{j(k-s)}(2^{kj+l})^{\sigma}.$$

This series can be summed exactly when $\mathrm{Re}(\sigma) < \sigma_0 = -(1 - s/k)$:

$$\hat{\omega}_l(\sigma) = 2^{l\sigma}\,\frac{2^{k-s+k\sigma}}{1 - 2^{k-s+k\sigma}}, \tag{55}$$

and expansion (54) shows that the differences $\omega_l(\sigma) - \hat{\omega}_l(\sigma)$ and $\omega_l'(\sigma) - \hat{\omega}_l(\sigma)$ have a general term of the form

$$O(2^{j(k\,\mathrm{Re}(\sigma)-s)}),$$

and therefore are analytic for $\mathrm{Re}(\sigma) < s/k$. Equation (54) also shows that

$$\omega_l(\sigma) - \hat{\omega}_l(\sigma) = O(|\sigma|^2), \qquad \omega_l'(\sigma) - \hat{\omega}_l(\sigma) = O(|\sigma|^2), \tag{56}$$

for large $|\sigma|$ with $-1 \leq \mathrm{Re}(\sigma) \leq s/2k$.

Thus,

$$\begin{aligned}\phi^*(\sigma) = {} & -\Gamma(\sigma)(1+\sigma)\sum_{l=0}^{k-1} \delta_1\, \delta_2 \cdots \delta_l \hat{\omega}_l(\sigma) \\ & - \Gamma(\sigma)\sum_{l=0}^{k-1} \delta_1\, \delta_2 \cdots \delta_l[\omega_l(\sigma) - \hat{\omega}_l(\sigma) + \sigma(\omega_l'(\sigma) - \hat{\omega}_l(\sigma))]\end{aligned}$$

and using (55) and (56) concludes the proof of the proposition. □

The final stage of the asymptotic analysis of $\phi(x)$ for large x, and thus of the asymptotics of $c_{u,n}$, is to use the inversion theorem for Mellin transforms,

$$\phi(x) = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} \phi^*(\sigma)x^{-\sigma}\, d\sigma, \qquad -1 < c < -\left(1 - \frac{s}{k}\right), \tag{57}$$

and, under suitable conditions, evaluate the integral using Cauchy's theorem as a sum of residues to the right of the vertical line $\{c + it \mid t \in \mathbb{R}\}$ and a remainder term of a small order when x is large.

We consider the integral

$$\phi_N(x) = \frac{1}{2i\pi}\int_{\Gamma_N} \phi^*(\sigma)x^{-\sigma}\, d\sigma, \tag{58}$$

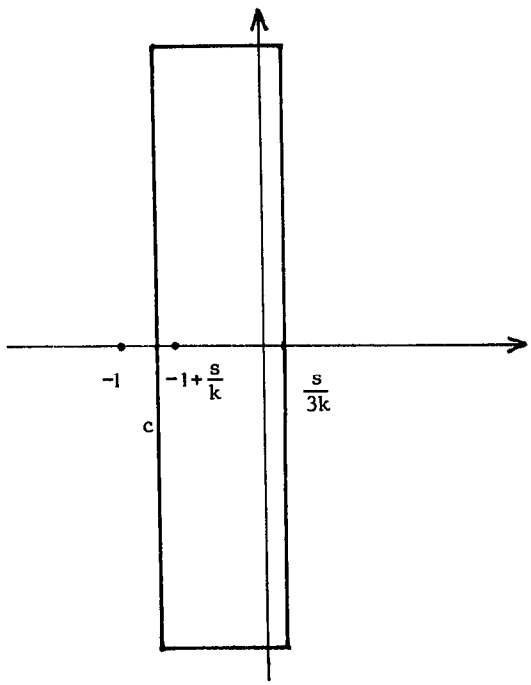


FIG. 3. The rectangular contour Γ_N used in evaluating $\phi(x)$ through an inverse Mellin transform.

where Γ_N is the rectangular contour oriented clockwise (see Figure 3):

$$\Gamma_N = \Gamma_N^1 + \Gamma_N^2 + \Gamma_N^3 + \Gamma_N^4,$$

$$\Gamma_N^1 = \left\{ c + it : |t| \le \frac{(2N+1)\pi}{k \log 2} \right\},$$

$$\Gamma_N^2 = \left\{ u + \frac{(2N+1)i\pi}{k \log 2} : c \le u \le \frac{s}{3k} \right\},$$

$$\Gamma_N^3 = \left\{ \frac{s}{3k} + it : |t| \le \frac{(2N+1)\pi}{k \log 2} \right\},$$

$$\Gamma_N^4 = \left\{ u - \frac{(2N+1)i\pi}{k \log 2} : c \le u \le \frac{s}{3k} \right\},$$

with N an integer (contours of a similar type are used for instance in [13, p. 132]).

Setting

$$\phi_N(x) = \phi_N^1(x) + \phi_N^2(x) + \phi_N^3(x) + \phi_N^4(x),$$

where ϕ_N^j corresponds to the contributions to integral (58) of the part Γ_N^j of the contour, we have the following results:

(i) $\phi_N^1(x) \to \phi(x)$ as $N \to \infty$,
(ii) $\phi_N^2(x) = o(1)$ as $N \to \infty$,
(iii) $|\phi_N^3(x)| \le x^{-s/(3k)} \int_{\Gamma_\infty} |\phi^*(\sigma)|\, d\sigma = O(x^{-s/(3k)})$,
(iv) $\phi_N^4(x) = o(1)$ as $N \to \infty$.

Of these assertions (i) is obvious by continuity; (ii) and (iv) come from the exponential decrease of $\Gamma(s)$ toward $i\infty$; (iii) is the trivial majorization of the absolute value of an integral.

Thus, letting N tend to infinity, we find

$$\phi_\infty(x) = \phi(x) + O(x^{-s/(3k)}). \tag{59}$$

Now the integral (58) can also be evaluated as the sum of the residues of the integrand inside Γ_N. As $N = \infty$, this sum is absolutely convergent and we have

$$\phi_\infty(x) = \underset{\alpha \in \mathrm{Pole}(\phi^*(\sigma))}{-\sum} \mathrm{Res}(\phi^*(\sigma)x^{-\sigma},\ \sigma = \alpha). \tag{60}$$

The poles of $\phi^*(\sigma)$ inside Γ_∞ are

—simple poles at

$$\alpha_j = \sigma_0 + \frac{2ij\pi}{k \log 2};$$

—a simple pole at $\sigma = 0$.

Thus, (59) and (60) can be rewritten as

$$\sigma(x) = \underset{\alpha \in \mathrm{Pole}(\phi^*(\sigma))}{-\sum} x^{-\alpha}\mathrm{Res}(\phi^*(\sigma),\ \sigma = \alpha) + O(x^{-s/(3k)}),$$

which is precisely an asymptotic expansion of $\phi(x)$ for large x. The contribution of the pole $\alpha = 0$ is $O(1)$; the contribution of $\alpha = \sigma_0$ is derived from the result of Proposition 2, and is

$$x^{-\sigma_0} \frac{(1 + \sigma_0)\Gamma(\sigma_0)}{k \log 2} \sum_{l=0}^{k-1} \delta_1\, \delta_2 \cdots \delta_l 2^{l\sigma_0}. \tag{61}$$

The contribution of α_j is similarly

$$x^{-\sigma_0}\exp\left(-\frac{2ij\pi}{k}\log_2 x\right)(1 + \alpha_j)\Gamma(\alpha_j) \sum_{l=0}^{k-1} \delta_1\, \delta_2 \cdots \delta_l 2^{l\alpha_j}, \tag{62}$$

so that

$$c_{u,n} = n^{1-s/k}\gamma\left(\frac{\log_2 n}{k}\right) + O(1),$$

with $\gamma(u)$ a periodic function of u, with period 1, mean value, and Fourier coefficients obtained from (61) and (62), respectively.

4.2 GRID-FILE ALGORITHMS. *Grid-file* or *extendible-cell* methods are a class of algorithms suitable for maintaining large collections of multiattribute records on secondary storage (see [16], [17], and [21]).

They are based on a dynamically varying partitioning of the underlying record space that adapts itself gracefully to the particular structure of the file being operated on. These algorithms can be viewed as multidimensional generalizations of dynamic hashing [14], extendible hashing [5], or virtual hashing [15].

If a suitable splitting policy is used (as in [21] or [17] when one uses level alternation for attribute splittings instead of time alternation), the paging of the file is equivalent to the paging of a k-d-trie.

Definition. The *paged k-d-trie* with page capacity b built on a file F is obtained from the k-d-tree built on F by placing in single pages all maximal subtrees containing at most b records.

The part of the tree obtained by pruning all leaf pages is called the *index* (or *directory*) of the paged k-d-trie.

This definition is illustrated by Figure 4.

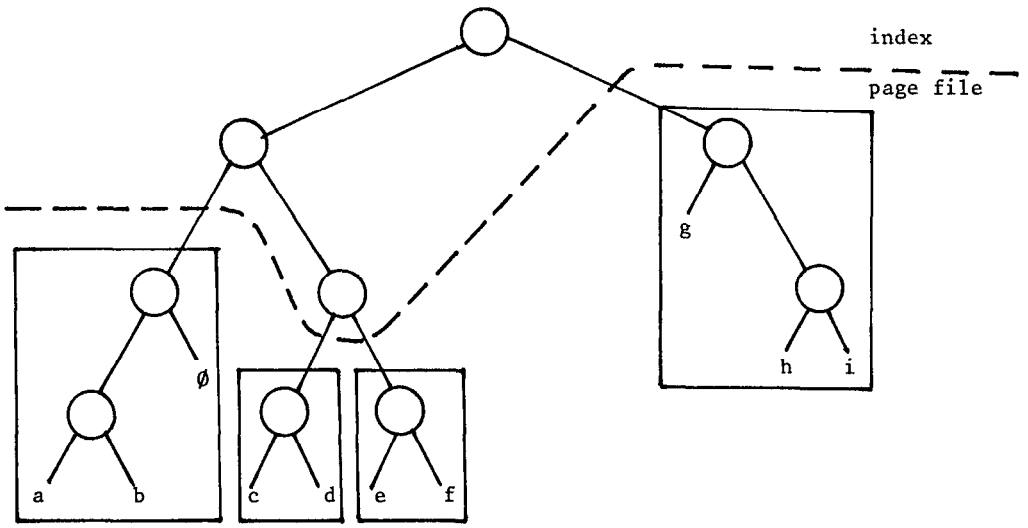


FIG. 4. The paged 2-d-trie corresponding to $F = \{a, b, \ldots, i\}$ with page capacity 3 when a = (00–, 00–), b = (00–, 01–), c = (00–, 10–), d = (00–, 11–), e = (01–, 10), f = (01–, 11–), g = (1–, 0–), h = (10–, 1–), and i = (11–, 1–).

The various schemes mentioned above differ by the way the index is implemented: it may be kept in core (as in dynamic hashing of [14]) or it may be represented as a perfect tree, embedding encoded into an array ([21] generalizing [5]), or as a multidimensional array [17].

The characteristic parameter of the cost of a partial match retrieval that is independent of the particular representaton of the index is the number of accesses to the paged file. Its expected value is given by the following theorem.

THEOREM 3. *The expected cost of a partial match query measured by the number of page accesses in a paged k-d-trie constructed from a file of size n (under the Bernoulli model) or from a file of expected size n (under the Poisson model) satisfies*

$$c_{u,n} = \gamma\left(\frac{1}{k}\log_2 n\right)n^{1-s/k} + O(1),$$

where $\gamma(u)$ is a periodic function of u with period 1 and mean value

$$\gamma_0 = \frac{-\Gamma(s/k-1)\binom{s/k+b-1}{b}}{k \log 2}$$
$$\times [(\delta_1 - 1) + \delta_1(\delta_2 - 1)2^{s/k-1} + \cdots + \delta_1\,\delta_2 \cdots \delta_{k-1}(\delta_k - 1)2^{(k-1)(s/k-1)}].$$

PROOF. The expected cost measured in the number of page accesses corresponding to a tree $t = t_1 \hat{} t_2$ for a random query with specification pattern u satisfies

$$c_u[t] = \begin{cases} \frac{\delta_1}{2}(c'_u[t_1] + c'_u[t_2]) & \text{if } |t| > b, \\ 1 & \text{if } |t| \le b. \end{cases}$$

It proves necessary for our later treatment to operate with the modified quantity

$$\hat{c}_u[t] = c_u[t] - 1,$$

which satisfies for all t the recurrence

$$\hat{c}_u[t] = \frac{\delta_1}{2}(\hat{c}_{u'}[t_1] + \hat{c}_{u'}[t]) + (\delta_1 - 1)\chi(|t| > b), \tag{63}$$

with $\chi(P)$ the characteristic function of predicate P.

The expected value $\hat{c}_{u,n} = c_{u,n} - 1$ of $\hat{c}_u[t]$ taken over all trees of size n under the Bernoulli model therefore satisfies

$$\hat{c}_{u,n} = \delta_1 \sum_{k=0}^{n} \frac{1}{2^n}\binom{n}{k}\hat{c}_{u',k} + (\delta_1 - 1)\chi(n > b). \tag{64}$$

The exponential generating function of the $c_{u,n}$,

$$\hat{c}_u(z) = \sum_{n\geq 0} \hat{c}_{u,n}\frac{z^n}{n!},$$

satisfies a relation obtained from (64),

$$\hat{c}_u(z) = \delta_1 e^{z/2}\hat{c}_{u'}\left(\frac{z}{2}\right) + (\delta_1 - 1)[e^z - e_b(z)], \tag{65}$$

with $e_b(z)$ denoting the truncated exponential

$$e_b(z) = \sum_{j=0}^{b}\frac{z^j}{j!}. \tag{66}$$

We now let $d_u(z)$ denote the quantity

$$d_u(z) = e^{-z}\hat{c}_u(z).$$

Thus, $d_u(z)$ is the expectation of $\hat{c}_u[t]$ if the number of elements in the file follows a Poisson distribution with parameter z. Equation (65) then leads to a difference system relating $d_u(z)$, $d_{u'}(z)$, ...:

$$\begin{aligned}
d_u(z) &= \delta_1 d_{u'}\left(\frac{z}{2}\right) + (\delta_1 - 1)(1 - e^{-z}e_b(z)),\\
d_{u'}(z) &= \delta_2 d_{u''}\left(\frac{z}{2}\right) + (\delta_2 - 1)(1 - e^{-z}e_b(z)),\\
&\vdots\\
d_{u^{(k-1)}}(z) &= \delta_k d_u\left(\frac{z}{2}\right) + (\delta_k - 1)(1 - e^{-z}e_b(z)).
\end{aligned} \tag{Σ}$$

From the combinatorial origin of parameters, we know that $d_u(z) = O(z^{b+1})$ for small z and $d_u(z) = O(z)$ for large z.

Thus, the Mellin transforms of $d_u(z)$, $d_{u'}(z)$, ... are all defined in the stripe

$$-(b+1) < \mathrm{Re}(\sigma) < -1. \tag{67}$$

We let now $d_u^*(\sigma)$ denote the Mellin transform of $d_u(z)$. From functional properties of the Mellin transforms recalled in the previous section follows that the transforms satisfy the linear system

$$\begin{aligned}
d_u^*(\sigma) &= \delta_1 2^\sigma d_u^*(\sigma) + (\delta_1 - 1)\alpha(\sigma),\\
d_{u'}(\sigma) &= \delta_2 2^\sigma d_{u''}^*(\sigma) + (\delta_2 - 1)\alpha(\sigma),\\
&\vdots\\
d_{u^{(k-1)}}^*(\sigma) &= \delta_k 2^\sigma d_u^*(\sigma) + (\delta_k - 1)\alpha(\sigma),
\end{aligned} \tag{Σ^*}$$

where

$$\alpha(\sigma) = \int_0^\infty (1 - e_b(x)e^{-x})x^{\sigma-1}\,dx.$$

This last transform is also defined in the stripe (67), and it can be computed by linearity. We find

$$\alpha(\sigma) = -\sum_{j=0}^{b} \frac{\Gamma(\sigma + j)}{j!} = -\Gamma(\sigma)\beta_b(\sigma),$$

where

$$\beta_b(\sigma) = 1 + \frac{\sigma}{1!} + \frac{\sigma(\sigma+1)}{2!} + \cdots + \frac{\sigma(\sigma+1)\cdots(\sigma+b-1)}{b!} = \binom{\sigma+b}{b}. \quad (68)$$

The system Σ^* can be solved in a manner similar to what was done before for k-d-tries; successively eliminating $d^*_{u'}$, $d^*_{u''}$, ..., we get

$$d^*_u(\sigma) = 2^{k-s}2^{k\sigma}d^*_u(\sigma) - \beta_b(\sigma)\Gamma(\sigma)\omega(\sigma),$$

whence, by solving for $d^*_u(\sigma)$,

$$d^*_u(\sigma) = \frac{-\beta_b(\sigma)\Gamma(\sigma)\omega(\sigma)}{1 - 2^{k(\sigma-\sigma_0)}}, \quad (69)$$

with

$$\begin{aligned}\omega(\sigma) = {} & (\delta_1 - 1) + \delta_1(\delta_2 - 1)2^\sigma \\ & + \delta_1\delta_2(\delta_3 - 1)2^{2\sigma} + \cdots + \delta_1\delta_2\cdots\delta_{k-1}(\delta_k - 1)2^{(k-1)\sigma}\end{aligned} \quad (70)$$

and

$$\sigma_0 = -\left(1 - \frac{s}{k}\right).$$

We can now conclude with the asymptotic analysis recovering $d_u(z)$ from $d^*_u(\sigma)$ by means of the inversion theorem for Mellin transforms, using the contours Γ_N of Section 4.1 and calculating residues as was done before. □

4.3 Other Analyses for Digital Structures. It should be clear by now that other cost measures (provided they are additive) of partial match can be analyzed in exactly the same way. Let us take as an example the number N_u of empty pages visited in the case of a search in a paged k-d-trie with page capacity b: Some implementations might skip the loading of such pages. N_u satisfies the recursive definition

$$N_u[t] = \begin{cases} \frac{\delta_1}{2}(N_{u'}[t_1] + N_{u'}[t_2]) & \text{if } b < |t|, \\ 0 & \text{if } 1 \le |t| \le b, \\ 1 & \text{if } 0 = |t|. \end{cases}$$

Thus, going to exponential generating functions of expected values $N_{u,n}$,

$$N_u(z) = \delta_1 e^{z/2}N_{u'}\left(\frac{z}{2}\right) + 1 - \delta_1 e_b\left(\frac{z}{2}\right).$$

Consider the modified quantities $\hat{N}_{u,n} = N_{u,n} - \delta_{0,n}$; then

$$\hat{N}_u(z) = \delta_1 e^{z/2}\hat{N}_{u'}\left(\frac{z}{2}\right) + \delta_1\left(e^{z/2} - e_b\left(\frac{z}{2}\right)\right).$$

The Mellin transform of $e^{-z}N_u(z)$ is the function $N_u^*(\sigma)$ which satisfies

$$N_u^*(\sigma) = \delta_1 2^\sigma N_{u'}^*(\sigma) + \delta_1 A(\sigma)\Gamma(\sigma),$$

where

$$A(\sigma) = 2\sigma - \sum_{j=0}^{b} \frac{\sigma(\sigma+1)\cdots(\sigma+j-1)}{j!}\frac{1}{2^j},$$

whence

$$N_u^*(\sigma) = \frac{A(\sigma)\bar{\omega}(\sigma)\Gamma(\sigma)}{1 - 2^{k(\sigma+1-s/k)}},$$

with

$$\bar{\omega}(\sigma) = \delta_1 + \delta_1\delta_2 2^\sigma + \cdots + \delta_1\,\delta_2\cdots\delta_k 2^{(k-1)\sigma}.$$

Hence, via a residue calculation and leaving aside periodic fluctuation terms, we have the approximate formula:

$$\frac{1}{n} N_{u,n} \approx \frac{A(s/k-1)\Gamma(s/k-1)\bar{\omega}(s/k-1)}{k \log 2}.$$

Notice that $A(\sigma)$ is the difference between $(1-x)^{-\sigma}$ and the first b terms of the Taylor expansion of $(1-x)^{-\sigma}$, at $x = \frac{1}{2}$. It thus becomes quite negligible if b is large enough. This is consistent with the fact that empty pages have a low probability of occurrence since the splitting of a b-page gives rise to an empty page with probability 2^{1-b} only. For instance, for $\sigma = \frac{1}{2}$, $|A(\sigma)|$ has values 0.0428, 0.0116, 0.005, and 7.10^{-6} when $b = 1, 2, 5$, and 10, respectively.

5. *Conclusions*

5.1 NUMERICAL VALUES. We have estimated numerically the expected costs of partial match queries for both k-d-trees and k-d-tries for $n = 500 \cdots 500{,}000$ and all partial match queries (PMQ) with dimension $k \le 4$.

In the case of k-d-trees, eliminating summations from recurrences permits determination of the $c_{u,n}$ in time $O(n)$. Since the forms of the asymptotic expansions are known by our Theorem 1, we can estimate from these exact values (say for $n \le 1000$) the coefficients in the first three terms of the asymptotic expansions and then use these values to estimate the $c_{u,n}$ for larger n (say until $n = 500{,}000$). Table I describes the results obtained in this way: Results for $n = 500$ are exact; results for $n = 5000, \cdots$ are obtained by such an extrapolation process. Experiments with exact values determined for $n = 5000$, 10,000 suggest that the accuracy of the results is ± 2 percent in all cases.

In the case of k-d-tries, the task is simpler since we have at our disposal the (exact) expansions provided by Lemma 8. The corresponding results are displayed in Table II.

In Tables I and II we have also indicated, in the third column, the dominant terms in the asymptotic expansions of the $c_{u,n}$ (leaving aside the fluctuating periodic terms in the case of tries). These dominant terms lead to values that are at most 5 percent off the exact values as soon as $n \ge 500$, and thus they provide a useful basis for comparisons. Notice also that the case of dimension $k = 2$ is covered in these tables by patterns $*S*S$ and $S*S*$.

One can observe that the periodicities, in the case of k-d-tries, appear quite distinct by comparing the costs of patterns $*SS*$ and $S**S$: The former leads to a

TABLE I. ESTIMATES OF THE COST OF A PARTIAL-MATCH QUERY IN A k-d-TREE FOR ALL SPECIFICATION PATTERNS WHEN DIMENSION $k \leq 4$

s/k	Pattern	Asymptote	500	5,000	50,000	500,000
1/4	$***S$	$1.66n^{0.78}$	224	1,392	8,593	52,988
	$**S*$	$1.49n^{0.78}$	201	1,246	7,691	47,423
	$*S**$	$1.33n^{0.78}$	180	1,115	6,883	42,443
	$S***$	$1.19n^{0.78}$	161	998	6,160	37,985
1/3	$**S$	$1.88n^{0.71}$	159	839	4,374	22,763
	$*S*$	$1.61n^{0.71}$	137	720	3,753	19,533
	$S**$	$1.38n^{0.71}$	117	618	3,221	16,762
2/4	$**SS$	$2.91n^{0.56}$	92	345	1,266	4,625
	$*S*S$	$2.55n^{0.56}$	80	302	1,108	4,045
	$*SS*$	$2.27n^{0.56}$	72	269	989	3,611
	$S**S$	$2.27n^{0.56}$	72	269	989	3,611
	$S*S*$	$1.99n^{0.56}$	63	236	865	3,158
	$SS**$	$1.77n^{0.56}$	56	210	772	2,820
2/3	$*SS$	$4.30n^{0.39}$	45	119	303	761
	$S*S$	$3.59n^{0.39}$	37	99	253	635
	$SS*$	$3.00n^{0.39}$	31	83	212	531
3/4	$*SSS$	$6.16n^{0.30}$	34	76	161	332
	$S*SS$	$5.34n^{0.30}$	29	66	139	288
	$SS*S$	$4.63n^{0.30}$	25	57	121	250
	$SSS*$	$4.02n^{0.30}$	22	50	105	217

TABLE II. EXACT VALUES OF THE COST OF A PARTIAL-MATCH QUERY IN A k-d-TRIE FOR ALL SPECIFICATION PATTERNS WHEN DIMENSION $k \leq 4$

s/k	Pattern	Asymptote	500	5,000	50,000	500,000
1/4	$***S$	$2.30n^{0.75}$	240	1,378	7,783	43,438
	$**S*$	$1.93n^{0.75}$	201	1,140	6,495	36,816
	$*S**$	$1.62n^{0.75}$	171	959	5,385	30,545
	$S***$	$1.36n^{0.75}$	145	819	4,567	25,473
1/3	$**S$	$2.47n^{0.66}$	154	724	3,368	15,607
	$*S*$	$1.96n^{0.66}$	122	572	2,658	12,352
	$S**$	$1.56n^{0.66}$	96	454	2,120	9,864
2/4	$**SS$	$3.72n^{0.50}$	78	260	840	2,654
	$*S*S$	$3.08n^{0.50}$	66	215	686	2,180
	$*SS*$	$2.63n^{0.50}$	56	181	585	1,880
	$S**S$	$2.63n^{0.50}$	56	186	588	1,841
	$S*S*$	$2.18n^{0.50}$	46	152	486	1,540
	$SS**$	$1.86n^{0.50}$	40	129	409	1,303
2/3	$*SS$	$5.01n^{0.33}$	34	80	179	392
	$S*S$	$3.97n^{0.33}$	27	64	142	312
	$SS*$	$3.15n^{0.33}$	22	50	113	247
3/4	$*SSS$	$7.07n^{0.25}$	25	51	98	180
	$S*SS$	$5.89n^{0.25}$	21	43	82	150
	$SS*S$	$4.95n^{0.25}$	18	36	68	126
	$SSS*$	$4.16n^{0.25}$	15	30	58	107

smaller cost for $n = 500$, 5000, 50,000, but to a larger cost for $n = 500{,}000$. Nevertheless, the amplitude of periodic fluctuations when $k \leq 4$ is limited to a few percent.

Tables I and II also agree with our expectation that "less specified" patterns have larger costs (in terms of exponents and/or multiplicative coefficients).

5.2 DIGITAL TRIES VERSUS SEARCH TREES. As already pointed out, the results from our analysis show that digital tries will always, for large enough n, dominate (multidimensional) search trees in terms of performances. In the case of k-d-tries, we have purposely taken as a cost measure the number of internal nodes traversed in order to have a homogeneous basis for comparison against k-d-trees. (Other cost measures could have been analyzed in exactly the same way, see Section 4.3.)

It appears from our data that, with a few minor exceptions, limited to $n = 500$, the cost of a query in a k-d-trie is always smaller than the corresponding cost for a k-d-tree. In the case of $n = 500{,}000$, the cost ratios vary from 1/1.2 (less specified patterns) to 1/2 (more specified patterns).

The fact that exponents differ is also confirmed by inspection of the ratios $c^{\text{trie}}_{u,n}/c^{\text{tree}}_{u,n}$. For $n = 500$ and $u = **S*$, the ratio is very nearly equal to 1/1 while for $n = 500{,}000$ it becomes 1/1.29. Similarly, for $u = SS**$ that ratio changes from 1/1.4 to 1/2.16 when n increases from 500 to 500,000.

We may observe at this stage that the better performance of k-d-tries should be related to the fact that digital tries tend to be better balanced than comparison-based search trees. For instance, the expected height of a random node in a random trie is $\log_2 n + O(1)$, whereas it is $\sim 2 \log n = 1.386 \log_2 n$ in a random search tree. According to that cost measure [13], tries are about 40 percent "better balanced."

Notice in this context that the cost of partial match query in a perfect k-d-tree (where all leaves are at the same level) has been shown by Bentley [1] to be

$$O(n^{1-s/k}).$$

Such shapes of trees may be obtained in the *static case* only (a fixed file) using preprocessing. Our analysis thus shows that k-d-tries lead to costs that are close (up to a multiplicative factor) to those of perfect trees, while k-d-trees generated by random insertions depart more significantly from that simple model. The function $\theta(s/k)$ that appears in the exponent of the cost of k-d-trees can thus be seen as reflecting the extra cost incurred by a dynamic usage of k-d-trees.

Finally, we should point out that the analyses we have presented here are relative to the cost of a *random* partial-match search in a *randomly* built tree. Since the probability that arguments to a PMQ coincide with some of the attributes contained in the tree is zero, that analysis reflects the cost of a *negative search.* Using exactly the same methods, we could equally well have studied the cost of a *positive search.* where specified attributes in the PMQ are conditioned to coincide with corresponding attributes of one of the elements of the file. The main conclusions, relative to the asymptotic orders of costs remain valid. An intuitive argument to support that fact is that a positive PMQ has probability about 1/2 of isolating an element to be found in the right subtree. Thus, with probability 1/2, a negative partial match in the left subtree, whose expected size is about $n/2$, will be necessary. Therefore, exponents should not depend on the positive or negative character of the query. Although, owing to the closeness of the analyses, we have not gone as far as computing the multiplicative constants involved, one should expect to observe the same type of dependency of the multiplicative constants with respect to the specification patterns as we have witnessed above.

5.3 METHODOLOGY. The analysis of k-d-tries is yet another illustration of the use of difference equations and Mellin transform techniques. The latter method is due to Knuth and de Bruijn (see [13, p. 131 ff], and also [9] and [10]). Notice, however, a stylistic variation of our Mellin transform analysis: Instead of establishing an *exponential approximation*, here somewhat clumsy, we have, in the analysis

of k-d-tries, transformed directly the exact expressions of average values, an approach that replaced exponential approximations by the analytic continuation result of Proposition 3.

The analysis of k-d-trees, via singularities of differential systems, uses what we feel to be novel techniques of some generality. Consider a *recursive splitting process* in which a set of n elements is partitioned (recursively) into a "left" subset (L) and a "right" subset (R) satisfying $|L| + |R| = n - 1$ (one element, the "root" is put aside) in such a way that the probability $\Pr(|L| = k)$ is a rational function of n and k. For a large class of such processes, the analysis of *additive costs* will lead to generating functions for average values that satisfy linear integral equations or, equivalently, to a linear system of differential equations. There, a singularity analysis like the one we have developed will permit to derivation of asymptotic estimates for the expected costs.

Such splitting processes occur for instance in relation to the analysis of quicksort, standard binary search trees, paged binary search trees, median-of-three quicksort, etc.

5.4 Differential Systems and Multiplicative Factors. The reader may have noticed that Theorems 2 and 3 do not provide *explicit values* for the multiplicative constants involved in the analysis of k-d-trees. Actually, it was necessary to resort to a nonconstructive argument in order to establish that, in all generality, the coefficients γ_u are nonzero.

This is not to be interpreted as a weakness of the method. The problem is comparable to finding, at some point, the value of a function satisfying a differential equation for which no closed-form solution exists. In both cases *numerical schemes* make it possible to determine these values to an arbitrary degree of accuracy. It is only for reasons of computational simplicity that we have not been using them here. (The method used above in Section 5.1 appeared to be reliable enough for all practical purposes.) However, in view of further applications of our methods, we informally indicate the principles of a computational procedure that may be used to determine the involved constants.

Assume first that Φ is a linear differential operator that is singular at $z = 1$ only, and, with $a(z)$ a known function, consider the solution to the equation

$$\Phi(f(z)) = a(z). \tag{NH}$$

with a set of initial conditions IC[0] on $f(z)$.

(a) Determine by the method of indeterminate coefficients (using the expansion of $a(z)$ around $z = 1$, $z = 0$, or some other point), a particular solution $g(z)$ to the equation

$$\Phi(g(z)) = a(z).$$

(b) If the system has a regular singular point and no logarithmic term occurs from confluence of solutions to the indicial equation, the general solution of the nonhomogeneous system (NH) is of the form

$$f(z) = g(z) + \sum_{\alpha} \frac{h_\alpha(z)}{(1 - z)^\alpha}.$$

Determine the Taylor expansions of the $h_\alpha(z)$ around $z = 1$ again using the method of indeterminate coefficients. These expansions depend only on the values of the $h_\alpha(1)$ that can be chosen arbitrarily.

(c) Identify the values of the $h_\alpha(1)$ for the particular solution $f(z)$ sought, using the *initial conditions* IC[0] on $f(z)$ at $z = 0$. Note that, the system being only singular at $z = 1$, the $h_\alpha(z)$ are entire functions. The numerical accuracy of the method is thus a function of the number of terms that are kept in the expansions of the $h_\alpha(z)$ around $z = 1$.

In our case, a further complication occurs since the system is also singular at $z = 0$, so that the expansions of the $h_\alpha(z)$ are not guaranteed to converge at $z = 0$. In that situation, replace step (c) above by the following sequence:

(1) Compute enough Taylor coefficients of the Taylor expansion of $f(z)$ around $z = 0$ (using initial values and the method of indeterminate coefficients).
(2) Use preceding values to determine with sufficient accuracy the value of $f(z)$ and its derivatives at $z = \frac{1}{2}$. This gives a new set of initial conditions IC[$\frac{1}{2}$].
(3) Identify the values $h_\alpha(1)$ from the new set of initial conditions IC[$\frac{1}{2}$]. (This corresponds to step (c) in the previous case.)

The method just sketched may be adapted to cope with more general situations where logarithmic terms appear. Combined with results derived from contour integration, which we have been using in this paper, it provides a numerical scheme that may be used to determine the multiplicative factors involved in asymptotic expansions of coefficients of a large class of solutions of differential systems. It permits, therefore, a complete asymptotic analysis for solutions of a large class of recurrences, as may be encountered in the field of analysis of algorithms.

ACKNOWLEDGMENTS. P. Flajolet would like to express his gratitude to Pr. Narasimhan, M. Joseph, and R. K. Shyamasundar for their invitation to visit the Tata Institute of Fundamental Research in Bombay, and to S. Joshi for several discussions that led to this work. The authors would like to thank K. Melhorn for stimulating discussions and the referee for a very careful scrutiny of the paper.

REFERENCES

1. BENTLEY, J. L. Multidimensional binary search trees used for associative searching. *Commun. ACM 18*, 9 (1975), 509–517.
2. CARDENAS, A. F., AND SAGAMANG, J. P. Doubly-chained tree data base organization—Analysis and design strategies. *Comput. J. 20* (1977), 15–26.
3. DAVIES, B. *Integral Transforms and Their Applications.* Springer-Verlag, New York, 1978.
4. DOETSCH, G. *Handbuch der Laplace transformation.* band II: *Anwendungen der Laplace transformation.* 1. *Abteilung.* Birkhaueser Verlag, Basel and Stuttgart, 1955.
5. FAGIN, R., NIEVERGELT, J., PIPPENGER, N., AND STRONG, H. R. Extendible hashing—A fast access method for dynamic files. *ACM Trans. Database Syst. 4*, 3 (1979), 315–344.
6. FINKEL, R. A., AND BENTLEY, J. L. Quad trees: A data structure for retrieval on composite keys. *Acta Inf. 4* (1974), 1–9.
7. FLAJOLET, P., AND ODLYZKO, A. The average height of binary trees and other simple trees. *J. Comput. Syst. Sci. 25*, 2 (1982), 171–213.
8. FLAJOLET, P., AND PUECH, C. Tree structures for partial match retrieval. In *Proceedings of the 24th Annual IEEE Symposium on the Foundations of Computer Science* (Tucson, Ariz.). IEEE, New York, 1983, pp. 282–288.
9. FLAJOLET, P., REGNIER, M., AND SEDGEWICK, R. Some uses of the Mellin integral transform in the analysis of algorithms. In *Proceedings of the NATO Advanced Study Institute on Combinatorial Algorithms on Words* (Maratea, Italy), NATO ASI Series, vol. F12. Springer-Verlag, 1985, pp. 241–254.
10. FLAJOLET, P., REGNIER, M., AND SOTTEAU, D. Algebraic methods for trie statistics. *Ann. Discrete Math. 25* (1985), 145–188.
11. HENRICI, P. *Applied and Computational Complex Analysis,* vol. 2. Wiley, New York, 1977.
12. KASHYAP, R. L., SUBAS, S. K. C., AND YAO, S. B. Analysis of the multiple-attribute-tree data base organization. *IEEE Trans. Softw. Eng. SE-3*, 6 (1977), 451–467.

13. KNUTH, D. E. *The Art of Computer Programming*, vol. 3. Addison-Wesley, Reading, Mass., 1973.
14. LARSON, P. A. Dynamic hashing. *BIT 18* (1978), 184–201.
15. LITWIN, W. Virtual hashing: A dynamically changing hashing. In *Proceedings of the 4th Conference on Very Large Data Bases* (Berlin). ACM, New York, 1978, pp. 517–522.
16. LLOYD, J. W., AND RAMAMOHANARAO, K. Partial-match retrieval for dynamic files. *BIT 22* (1982), 150–168.
17. NIEVERGELT, J., HINTERBERGER, H., AND SEVCIK, K. C. The grid file: An adaptable, symmetric multikey file structure. *ACM Trans. Database Syst. 9*, 1 (Mar. 1984), 38–71.
18. ODLYZKO, A. Periodic oscillations of coefficients of power series that satisfy functional equations. *Adv. Math 44* (1982), 180–205.
19. REGNIER, M. Evaluation des performances du hachage dynamique. Thèse de 3ème cycle, Univ. de Paris-Sud, Orsay, France. April 1983.
20. RIVEST, R. L. Partial-match retrieval algorithms. *SIAM J. Comput. 5*, 1 (1976), 19–50.
21. TAMMINEN, M. The extendible cell method for closest point problems. *BIT 22* (1982), 27–41.

RECEIVED JULY 1983; REVISED FEBRUARY 1985; ACCEPTED AUGUST 1985