ELSEVIER
Theoretical Computer Science 144 (1995) 199–220
Theoretical Computer Science

Analysis of an optimized search algorithm for skip lists[☆]

Peter Kirschenhofer[a,1], Conrado Martínez[b,*], Helmut Prodinger[a,1]

[a] *Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstraße 8-10, A-1040 Vienna, Austria*

[b] *Departament de Llenguatges i Sistemes Informàtics, Polytechnical University of Catalonia, Pau Gargallo 5, E-08028 Barcelona, Spain*

Abstract

It was suggested in Pugh (1990) to avoid redundant key comparisons in the skip list search algorithm by marking those elements whose key has already been compared against the search key. We present here a precise analysis of the total search cost (expectation and variance), where the cost of the search is measured in terms of the number of key-to-key comparisons. These results are then compared with the corresponding values of the standard search algorithm.

1. Introduction

Skip lists have recently been introduced as a type of list-based data structure that may substitute search trees [10]. A set of n elements is stored in a collection of sorted linear linked lists in the following manner: all elements are stored in increasing order in a linked list called *level* 1 and, recursively, each element which appears in the linked list *level* i is included with independent probability q ($0 < q < 1$) in the linked list *level* $i + 1$.

The *level* of an element x is the number of linked lists it belongs to. For each element in the skip list, we need a node to store its key and as many pointers as its level indicates. The successor of x at the list level i is given by the ith pointer of x, also called ith *forward* pointer of x. A *header* points to the first element in each of the linked lists and it also stores the *height* of the skip list, which is the maximum level among the levels of the elements or total number of nonempty linked lists.

A detailed description of several skip list algorithms, as well as variants of the data structure, can be found in [9]. Interesting analytic aspects of the average-case

☆ This research was supported by the program Acciones Integradas Hispano-Austríacas (Austrian-Spanish Scientific Exchange Program). The second author was also supported by the ESPRIT BRA Program of the EC under contract no. 7141, project ALCOM II.

* Corresponding author. E-mail: conrado@lsi.upc.es

1 e-mail: {kirsch, proding}@rsmb.tuwien.ac.at.

SSDI 0304-3975(94)00296-7

performance of skip lists algorithms may be found in [1, 6, 7]. In [5] the probabilistic analysis of the search cost was considered in a slightly different way, namely, performing the asymptotic analysis of the *total search cost* or *path length*, i.e. the sum of the successful search costs to find all the elements in the data structure. In particular, the variance of this parameter was analyzed precisely.

The present paper is devoted to the analysis of the total search cost of an *optimized version* of the search algorithm that greatly reduces the number of key-to-key comparisons. The optimized version of the search guarantees that the search key will be compared at most once with the key of any element in the skip list; additional pointer-to-pointer comparisons may be needed for those elements whose key has already been compared with the search key. If a key comparison is more expensive than a pointer comparison, this optimized version will be useful. It is also interesting to use this kind of algorithm when searching for elements in the skip list in parallel. If each process owns a variable `alreadyChecked` a large amount of concurrent accesses to the keys of the elements can be avoided. This technique has been used in the design of efficient parallel search and update algorithms for skip lists [3]. The optimized search algorithm was proposed in [9], together with an estimate of the average savings in the number of key-to-key comparisons. We give in the next section the precise asymptotic behavior of the average number of key comparisons and compare our result with Pugh's estimate.

According to the standard algorithm the search for an element is performed by traversing forward pointers as long as the key of the successor of the current node is smaller than the search key. When this traversal stops at the current level, the search is continued one level below. Clearly the algorithm terminates at level 1 when we are just one node in front of the node that contains the element we are looking for (we assume that the desired element is already in the skip list). In Fig. 1 we depict the search path for the 6th element in a skip list of size 10 and height 11.

In the optimized version of the search algorithm the number of key-to-key comparisons is reduced by assuring that the search key is never compared against the key of an element more than once. To this aim, the variable `alreadyChecked` is introduced. At the beginning this variable is set to "NIL". We then follow forward pointers as long as the elements pointed to are different from `alreadyChecked` and the keys of those elements are smaller than the search key. As soon as this horizontal traversal ends, `alreadyChecked` is set to the element pointed to at this moment and the search continues one level below (see Fig. 2).

Each dashed arrow and thick horizontal line in Fig. 1 corresponds to an "expensive" key comparison of the search key against the key of the item that is pointed to. The dotted arrows correspond to successful "cheap" pointer comparisons. For each of these successful pointer-to-pointer comparisons, the search path must drop one level. In our example we have 14 comparisons altogether. These split into 6 expensive key-to-key comparisons and 8 cheap successful pointer-to-pointer comparisons.

In order to analyze the total search cost, i.e. the total number of key comparisons made with this optimized algorithm to search each of the elements in a skip list, it is

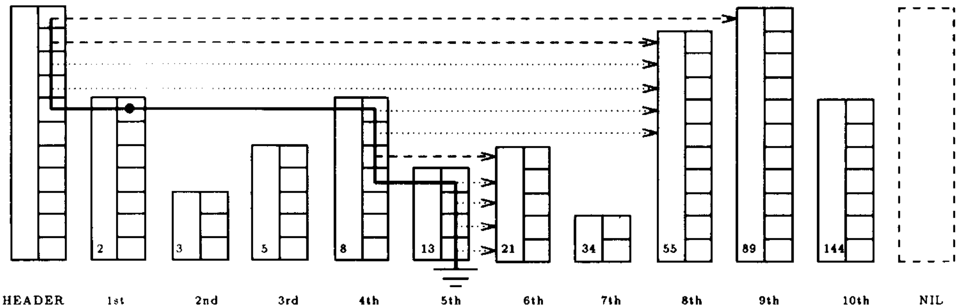

Fig. 1. An example of a search path in a skip list.

```
x := header(S); l := height(S);
alreadyChecked := NIL;
while l > 0 do
    {in the expression A cand B, B is evaluated only if A is true}
    while (x ↑ .forward[l] ≠ alreadyChecked) cand
            (x ↑ .forward[l] ↑ .key < search_key) do
        x := x ↑ .forward[l]
    end;
    alreadyChecked := x ↑ .forward[l];
    l := l − 1
end
```

Fig. 2. Optimized search algorithm (see [9]).

helpful to describe a skip list of size n as an n-tuple $(a_1, \ldots, a_n)$, where a_i denotes the level of the ith element. For instance, the skip list in Fig. 1 is described by the 10-tuple $(7, 3, 5, 7, 4, 5, 2, 9, 11, 8)$.

The probabilistic model for *random skip lists* describes a random skip list as the outcome of n independent identically distributed random variables. In particular, each $a_i \in \mathbb{N}$ is the outcome of a *geometric* random variable G_i of parameter p, i.e. $\text{Prob}\{G_i = k\} = pq^{k-1}$, where $q = 1 - p$. (Note that in some earlier papers the roles of p and q are interchanged.)

It is of interest to translate the search cost parameter into terms of *order statistics*. The number of key-to-key comparisons when searching for the ith element with the optimized algorithm can be split up into two contributions:

(1) Key comparisons where the search key is compared with a key that is larger or equal (there is a dashed horizontal arrow in Fig. 1 for each comparison of this kind).

(2) Key comparisons where the search key is compared with a key that is smaller (there is a thick horizontal line in Fig. 1 for each comparison of this kind).

It is an immediate observation that the comparisons of type (1) correspond bijectively to the strict left-to-right maxima of the sequence $(a_i, \ldots, a_n)$ whereas the comparisons of type (2) correspond bijectively to the weak right-to-left maxima of $(a_1, \ldots, a_{i-1})$.

To be precise, let us say that a_j $(i \leqslant j \leqslant n)$ is a strict left-to-right maximum (sLR, for short) of $(a_i, \ldots, a_n)$ if it is larger than $a_i, \ldots, a_{j-1}$, and that a_k $(1 \leqslant k \leqslant i-1)$ is a weak right-to-left maximum (wRL, for short) of $(a_1, \ldots, a_{i-1})$ if it is larger than or equal to $a_{k+1}, \ldots, a_{i-1}$.

Observe that for a fixed element i the two parameters are independent random variables; but this is no longer true for the total search cost, which is given by the sum of the number of sLR of all suffixes of $(a_1, \ldots, a_n)$ and of the number of wRL of all prefixes of $(a_1, \ldots, a_{n-1})$. This dependency is the reason that we cannot simply add the variances of these two parameters, which were already computed in [5, 8].

A second observation is the fact that the number of wRL of $(a_1, \ldots, a_n)$ is never counted above. This unpleasant asymmetry between the cumulation of sLR and wRL would lead to cumbersome recurrences in the probabilistic analysis. Therefore, we shift our attention to the *total unsuccessful search cost*. Let $C_{n,i}$ denote the cost of an unsuccessful search of a key belonging to the interval (x_{i-1}, x_i) in a random skip list of n elements, where x_i denotes the key of the ith element. By convention, $x_0 = -\infty$ and $x_{n+1} = +\infty$. Then, C_n, the total unsuccessful search cost is

$$C_n = \sum_{1 \leqslant i \leqslant n+1} C_{n,i}. \tag{1}$$

Obviously, $C_{n,i}$ is also the successful search cost for the ith element in a random skip list of n elements, for $i = 1, \ldots, n$. It turns out that C_n fulfills a nice recurrence relation that greatly simplifies the analysis. Furthermore, in Section 4, we will show that the two first moments of C_n are asymptotically equivalent to those of $\tilde{C}_n$, the *total successful search cost* in a random skip list of n elements.

From our previous discussion it follows that $C_{n,i} = \ell_{n,i} + r_i$, where $\ell_{n,i}$ is the number of sLR in $(a_i, \ldots, a_n)$ and r_i is the number of wRL in $(a_1, \ldots, a_{i-1})$. We assume $r_1 = 0$ and $\ell_{n,n+1} = 0$. The expectation and variance of these random variables is known [8]; moreover, they are independent and hence we have $E(C_{n,i}) = E(\ell_{n,i}) + E(r_i)$ and $\mathrm{Var}(C_{n,i}) = \mathrm{Var}(\ell_{n,i}) + \mathrm{Var}(r_i)$.

Theorem 1.1. *The expectations and variances of* $\ell_{n,i}$, *the number of* sLR *in* $(a_i, \ldots, a_n)$, *and* r_i, *the number of* wRL *in* $(a_1, \ldots, a_{i-1})$ *are*

$$E(\ell_{n,i}) = p\left[\log_Q(n-i+1) + \frac{\gamma}{L} + \frac{1}{2} - \frac{1}{L}\delta_1(\log_Q(n-i+1))\right] + \mathrm{O}\left(\frac{1}{n-i+1}\right), \quad n-i \to \infty,$$

$$E(r_i) = \frac{p}{q}\left[\log_Q(i-1) + \frac{\gamma}{L} - \frac{1}{2} - \frac{1}{L}\delta_1(\log_Q(i-1))\right] + \mathrm{O}\left(\frac{1}{i-1}\right), \quad i \to \infty,$$

$$\operatorname{Var}(\ell_{n,i}) = pq\log_Q(n-i+1) + p^2\left(-\frac{5}{12}+\frac{\pi^2}{6L^2}-\frac{\gamma}{L}-\frac{1}{L^2}[\delta_1^2]_0\right) + p\left(\frac{\gamma}{L}+\frac{1}{2}\right)$$

$$+\,\delta_2(\log_Q(n-i+1)) + \mathrm{O}\left(\frac{1}{n-i+1}\right), \quad n-i\to\infty,$$

$$\operatorname{Var}(r_i) = \frac{p}{q^2}\log_Q(i-1) + \frac{p^2}{q^2}\left(-\frac{5}{12}+\frac{\pi^2}{6L^2}+\frac{\gamma}{L}-\frac{2}{L}-\frac{1}{L^2}[\delta_1^2]_0\right) + \frac{p}{q}\left(\frac{\gamma}{L}-\frac{1}{2}\right)$$

$$+\,\delta_3(\log_Q(i-1)) + \mathrm{O}\left(\frac{1}{i-1}\right), \quad i\to\infty,$$

where $Q = q^{-1}$, $L = \log Q$, γ *is Euler's constant*, $\delta_1(x) = \sum_{k\neq 0}\Gamma(-2k\pi\mathrm{i}/L)\mathrm{e}^{2k\pi\mathrm{i}x}$ *is a periodic function of period* 1 *and mean* 0, $[\delta_1^2]_0$ *is the mean of the square of* δ_1 *and* δ_2 *and* δ_3 *are other periodic functions of period* 1 *and mean* 0.

Finally, there is a nice and suggestive interpretation for the unsuccessful search cost $C_{n,n+1}$ in terms of the skip list algorithm, which will be very helpful in the sequel: $C_{n,n+1} = r_{n+1}$ equals the number of key comparisons if we search for a key larger than any other key already in the skip list (or equivalently, if we search for "NIL", which is marked as alreadyChecked from the very beginning).

Now we are ready to start our analysis. The main part of the paper will be organized as follows.

In Section 2 we start from a combinatorial decomposition of random skip lists in order to get a functional equation for the probability generating function of the total unsuccessful search cost. This allows to compute the asymptotics of the expectation in a straightforward manner. In Section 3 we concentrate on our main result and evaluate the variance of C_n. In Section 4 we prove that our asymptotic results remain true for the total successful search cost $\tilde{C}_n$. Finally, in Section 5 we discuss some generalizations of the algebraic techniques used in the probabilistic analysis of skip list algorithms.

2. Probability generating functions and expectations

In order to derive a recurrence relation for the probability generating function of the total unsuccessful search cost, it is convenient to consider the following combinatorial decomposition of a skip list of height m (see Fig. 3): We split up the whole skip list $S = (a_1, \ldots, a_n)$ according to the leftmost appearance of an element $a_i = m$ into the partitioning element m and two skip lists $\sigma = (a_1, \ldots, a_{i-1})$ and $\tau = (a_{i+1}, \ldots, a_n)$. Observe that σ has height less than m and τ has height less than or equal to m; each of σ and τ have fewer elements than the skip list S.

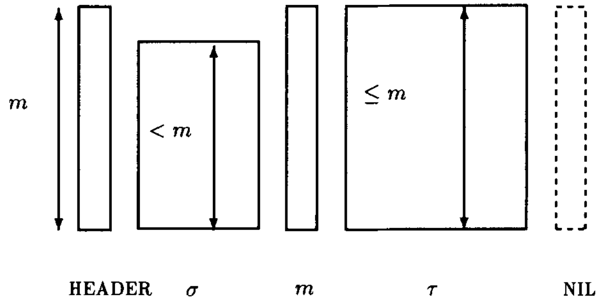

Fig. 3. Recursive decomposition of a skip list.

Following the previous discussion, the total unsuccessful search cost $C(S)^2$ of a skip list S is the sum of two contributions $L(S)$ and $R(S)$, where $L(S)$ is the cumulated number of sLR of all suffixes of S and $R(S)$ is the cumulated number of wRL of all prefixes of S.

Now, it is plain to see that

$$L(\sigma m \tau) = L(\sigma) + L(\tau) + |\sigma| + 1, \tag{2}$$

since each element at the left of $a_i = m$, i.e. in the σ-part, contributes as many sLR to $L(S)$ as it contributes to $L(\sigma)$ plus 1, the additional sLR corresponding to the partitioning element (the latter contribution is $|\sigma|$). The partitioning element itself contributes 1 to $L(S)$, and the elements in the τ-part contribute as many sLR to $L(S)$ as they do to $L(\tau)$.

For $R(S)$ we can argue as above. It is however easier to imagine – since we are summing up all these numbers – that, when considering some particular element, we are interested in the wRL to the left of it, the element itself being contributing as a wRL. Note that this approach takes into account the contribution of wRL of the whole sequence, corresponding to the unsuccessful search of a key larger than any other in the skip list, without the need of dealing with the "NIL" element.

Then, if an element is in the σ-part, it contributes as many wRL to $R(S)$ as it does to $R(\sigma)$. If it is the partitioning element $a_i = m$, the contribution is 1, and if it is in the τ-part, we must add 1 to the number of wRL it already contributed to the τ-part. This gives us a similar recursion for $R(S)$

$$R(\sigma m \tau) = R(\sigma) + R(\tau) + |\tau| + 1. \tag{3}$$

Altogether we have

$$C(\sigma m \tau) = C(\sigma) + C(\tau) + |S| + 1. \tag{4}$$

[2] We shall omit the explicit subindex n in $C(S)$, $L(S)$ and $R(S)$, since it is clear that $n = |S|$.

If we forget about the L(resp. R) contributions, the recurrence above may also be obtained by observing that the key of the partitioning element will always be compared with the search key no matter which element we are searching for.

Let us denote by $P^*(z, y)$ the bivariate generating function where the coefficient of $z^n y^k$ denotes the probability that a random skip list of size n has height fulfilling condition $(*)$ and the total unsuccessful search cost is equal to k. In general, for any generating function $f(z)$ over skip lists we shall denote $f^*(z)$ the corresponding generating function over skip lists whose height satisfies the condition $(*)$. Eq. (4) immediately translates to the functional equation

$$P^{=m}(z,y) = pq^{m-1} zy^2 P^{<m}(zy,y) P^{\leqslant m}(zy,y), \quad m \geqslant 1,$$
$$P^{=0}(z,y) = 1, \tag{5}$$

since the probability for a fixed element a_i to have value m is pq^{m-1}, and $|S| + 1 = |\sigma| + |\tau| + 2$, so that the contribution of the additional term in Eq. (4) splits up as

$$y^{|S|+1} = y^{|\sigma|} y^{|\tau|} y^2.$$

It is somehow easier to work with the generating functions $R^*(z,y) := zP^*(z,y)$, because the recursion reads now

$$R^{=m}(z,y) = pq^{m-1} R^{<m}(zy,y) R^{\leqslant m}(zy,y), \quad m \geqslant 1,$$
$$R^{=0}(z,y) = z. \tag{6}$$

Using the decomposition of the cost $C(S) = L(S) + R(S)$, the asymptotic behavior of the expectation would be a simple corollary of the results in [5], since we may add the expectations of L and R. However, to make the paper more self-contained, we give some details about the techniques and intermediate steps in the derivation of the asymptotic behavior of the expectation of C_n. It is also useful to present these computations here, since we shall apply similar techniques to compute the variance.

Let us introduce some handy abbreviations (the first two were already used in Theorem 1.1): $Q := q^{-1}$, $L := \log Q$, and

$$[\![m]\!] := 1 - z(1 - q^m).$$

Note that $[\![m]\!] = [\![m-1]\!] - pq^{m-1}z$.

We obtain the generating function $S^{=m}(z)$ of the expectations in the usual way by differentiating $R^{=m}(z,y)$ w.r.t. y and setting $y = 1$. Doing that, we find the recursion

$$S^{=m}(z) = \left.\frac{\partial R^{=m}}{\partial y}\right|_{y=1} = pq^{m-1}\left[\left(\frac{z}{[\![m-1]\!]^2} + S^{<m}(z)\right)\frac{z}{[\![m]\!]} + \left(\frac{z}{[\![m]\!]^2} + S^{\leqslant m}(z)\right)\frac{z}{[\![m-1]\!]}\right].$$

Since $S^{=m}(z) = S^{\leqslant m}(z) - S^{<m}(z)$, we can rewrite the equation above as

$$S^{\leqslant m}(z)[\![m]\!]^2 = S^{<m}(z)[\![m-1]\!]^2 + pq^{m-1}z^2\left(\frac{1}{[\![m]\!]} + \frac{1}{[\![m-1]\!]}\right), \tag{7}$$

and solve it by *iteration*,

$$S^{\leqslant m}(z) = \frac{p}{q}\frac{z^2}{[\![m]\!]^2}\sum_{i=1}^{m}\frac{q^i}{[\![i]\!]} + p\frac{z^2}{[\![m]\!]^2}\sum_{i=0}^{m-1}\frac{q^i}{[\![i]\!]}. \tag{8}$$

We are in fact interested in $S(z) := \lim_{m\to\infty} S^{\leqslant m}(z)$. Performing the limit for $m \to \infty$, we find

$$S(z) = \frac{p}{q}\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^i}{[\![i]\!]} + p\frac{z^2}{(1-z)^2}\sum_{i\geqslant 0}\frac{q^i}{[\![i]\!]}. \tag{9}$$

The expected value of interest is the coefficient of z^{n+1} in this expression; $[z^{n+1}]S(z) = E(C_n)$. It could be obtained for instance by *partial fraction decomposition*. However, for more complicated expressions as the ones we shall encounter in the next section, such an approach is not feasible and we will therefore use a more sophisticated procedure. It was already used in the previous paper [5], but we would like to present it again in a slightly rephrased form. The standard substitution

$$z = \frac{w}{w-1}$$

proves here to be very useful. In general,

$$[z^n]f(z) = (-1)^n[w^n](1-w)^{n-1}f(w/(w-1)).$$

This formula can be easily seen by (formal) residue calculus, as explained for instance in [4]. Our expressions will usually look nicer when expressed with the variable w, since

$$[\![i]\!] = \frac{1-wq^i}{1-w},$$

leading to expressions that belong to the class of the so-called *harmonic sums*

$$F(w) = \sum_i a_i f(b_i w),$$

where the coefficient of w^n in F satisfies

$$[w^n]\,F(w) = \sum_i a_i b_i^n \cdot [w^n]f(w).$$

Quite often, the series $\sum_i a_i b_i^n$ has a closed-form representation, and $[w^n]f(w)$ can be computed explicitly.

We would like to show this paradigm by considering

$$[z^n]\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^i}{[\![i]\!]}.$$

Following the method that we sketched above

$$\begin{aligned}[z^n]\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^i}{[\![i]\!]} &= (-1)^n[w^n]\,w(1-w)^n\sum_{i\geqslant 1}\frac{wq^i}{1-wq^i}\\ &= (-1)^n\sum_{k=1}^{n}\binom{n}{k}(-1)^{n-k}[w^{k-1}]\sum_{i\geqslant 1}\frac{wq^i}{1-wq^i}\\ &= \sum_{k=1}^{n}\binom{n}{k}(-1)^k\sum_{i\geqslant 1}q^{(k-1)i}[w^{k-1}]\frac{w}{1-w}\\ &= \sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{1}{Q^{k-1}-1}.\end{aligned}$$

This example is typical: the answer comes out as an alternating sum, involving both binomial coefficients and some "known" quantities. While such a form is not very convenient for numerical purposes because of the cancellations that occur, it is very handy for the asymptotic evaluation. Such a sum can be written as a Rice integral, and asymptotics are obtained simply by considering appropriate residues.

The survey paper [2] (in this issue) explains this methodology in detail.

Here, we confine ourselves to the basic formula

$$\sum_{k=a}^{n}\binom{n}{k}(-1)^k f(k) = -\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}} B(n+1,\,-z)\,f(z)\,\mathrm{d}z,$$

where $B(x,y)=\Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the classical Beta function, $\mathscr{C}$ is a positively oriented curve that encircles the points $a, a+1,\dots,n$ but not the points $0,1,\dots,a-1$ and $f(z)$ is analytic inside $\mathscr{C}$ and is a continuation to the complex plane of the sequence $f(k)$.

If $f(z)$ decreases sufficiently fast towards $\pm\mathrm{i}\infty$, one may perform an asymptotic evaluation by extending the contour of the integral to the left. The newly encountered negative residues will give the terms in the asymptotic expansion. See [11] for detailed analytical information.

In our instance, we find

$$E(C_n)=[z^{n+1}]S(z)=p(Q+1)\sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{1}{Q^{k-1}-1}+pn. \qquad (10)$$

Rice's method can be applied in this particular case using the obvious continuation $1/(Q^{z-1}-1)$ for the sequence $1/(Q^{k-1}-1)$.

The residue computations are standard and can be done by some computer algebra system, e.g. Maple, giving us the asymptotic behavior of $E(C_n)$.

Theorem 2.1 [Expected total unsuccessful search cost (optimized)]. *The expected value of total unsuccessful search cost C_n in a random skip list of n elements is, as $n \to \infty$,*

$$E(C_n) = (Q - q)n\log_Q n + n(Q - q)\left(\frac{(\gamma - 1)}{L} - \frac{1}{2} + \frac{1}{Q+1} + \frac{1}{L}\delta_4(\log_Q n)\right)$$

$$+ \mathrm{O}(\log n),$$

where $\delta_4(x) = \sum_{k \neq 0} \Gamma(-1 - 2k\pi i/L)\, e^{2k\pi i x}$ is a periodic function of period 1 and mean zero.

Since the total successful search cost satisfies $\tilde{C}_n = C_n - C_{n,n+1}$, we have

$$E(\tilde{C}_n) = E(C_n) + \mathrm{O}(\log n), \quad n \to \infty$$

because $E(C_{n,n+1}) = E(r_{n+1}) = \mathrm{O}(\log n)$.

We can then compare this last result with the previous result for the standard search algorithm.

Theorem 2.2 [Expected total successful search cost (standard)]. *The expected total successful search cost $\tilde{C}_n^{[\mathrm{st}]}$ in a random skip of n elements is, as $n \to \infty$,*

$$E(\tilde{C}_n^{[\mathrm{st}]}) = Qn\log_Q n + n\left(\frac{(\gamma - 1)Q + 1}{L} - \frac{Q}{2} + 1 + \frac{1}{L}\delta_5(\log_Q n)\right) + \mathrm{O}(\log n),$$

where $\delta_5 = \delta_4 - \delta_1$ is a periodic function of period 1 and mean zero [5].

Comparing the leading terms in both instances, we see that, asymptotically, we save about $qn\log_Q n$ key comparisons by using the optimized version. It is interesting to study the factor $q/\log Q$ as a function of q. The savings increase as $q \to 1$, but it should be clear that the total number of steps (including all types of comparisons, pointer inspections, ...) in both versions of the search are the same and that a larger value of q leads to a larger expected number of pointers per element. Therefore, it is not wise to choose a large q, but to look for a value of q that trades off key comparison savings, total number of comparisons and storage requirements. A plot of the coefficients K and K' of the $n\log n$ term in $E(C_n)$ (in $E(\tilde{C}_n)$, as a matter of fact) and $E(\tilde{C}_n^{[\mathrm{st}]})$ is given in Fig. 4. The degenerate case of $q = 1$ where K achieves a minimum reflects the hypothetical situation where all items are infinitely "tall" and hence infinitely many comparisons are just successful cheap pointer-to-pointer comparisons.

Pugh [9] claims that the optimized algorithm saves an average of $q(\log_Q n + q/p^2)$ comparisons per element. Pugh gets upper bounds for both the expected standard search cost and the expected optimized search cost. Then he gives his estimate subtracting these upper bounds. Subtracting upper bounds does not yield any firmly established conclusion. If the upper bounds are tight, one can assume that the difference should not be far away from the real value: this is indeed the case, since Pugh's estimate gives the right main asymptotic behavior for the average savings.

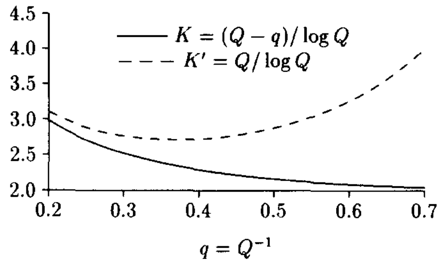

Fig. 4. Behavior of K and K' as a function of q.

3. Variance of the total unsuccessful search cost

In this section, we will compute the asymptotic behavior of the second factorial moment of C_n and of its variance. The generating function for the second factorial moment can be obtained differentiating $P^*(z,y)$ twice w.r.t. y and setting $y=1$. We shall proceed starting from $R^*(z,y)$ instead, because the recurrences for R^* are easier.

Let us recall Eq. (6)

$$R^{=m}(z,y) = pq^{m-1}R^{<m}(zy,y)R^{\leqslant m}(zy,y), \quad m \geqslant 1,$$
$$R^{=0}(z,y) = z.$$

Let $T^*(z) := R^*_{yy}(z,1)$, where a subscript x means partial derivative with respect to x. It is not difficult to see that

$$R^{\lessgtr m}(z,1) = \frac{z}{[\![m]\!]}, \qquad R_z^{\lessgtr m}(z,1) = \frac{1}{[\![m]\!]^2}, \qquad R_{zz}^{\lessgtr m}(z,1) = \frac{2(1-q^m)}{[\![m]\!]^3},$$
$$R_y^{\lessgtr m}(z,1) = S^{\lessgtr m}(z), \qquad R_{zy}^{\lessgtr m}(z,1) = S_z^{\lessgtr m}(z),$$

where $S^{\lessgtr m}$ was defined in Section 2.

Using the equalities above and collecting terms yields

$$\begin{aligned} &T^{\leqslant m}(z)[\![m]\!]^2 \\ &= T^{<m}(z)[\![m-1]\!]^2 + 2pq^{m-1}\left[\frac{z^2}{[\![m]\!][\![m-1]\!]} + \frac{z^3(1-q^m)}{[\![m]\!]^2} + \frac{z^3(1-q^{m-1})}{[\![m-1]\!]^2}\right. \\ &\quad + \frac{z}{[\![m-1]\!]}[\![m]\!]\,S^{\leqslant m}(z) + \frac{z}{[\![m]\!]}[\![m-1]\!]\,S^{<m}(z) + [\![m]\!]\,[\![m-1]\!]\,S^{\leqslant m}(z)S^{<m}(z) \\ &\quad \left. + z^2[\![m]\!]\,S_z^{\leqslant m}(z) + z^2[\![m-1]\!]\,S_z^{<m}(z)\right]. \end{aligned}$$

We can solve this recurrence by iteration, as we did for $S^{\lessgtr m}(z)$ in Section 2. Then we should compute the limiting generating function $T(z) := \lim_{m\to\infty} T^{\leqslant m}(z)$. Finally, dividing by z the generating function $T(z)$ gives us $H(z)$, the generating function for second factorial moments:

$$H(z) = 2\frac{p}{q}\frac{z}{(1-z)^2}\left((1+q)\sum_{i\geq 1} a_i(z)q^i + \sum_{i\geq 1} d_i(z)\,d_{i-1}(z)\,q^i\right), \tag{11}$$

where

$$a_i(z) = \frac{z(1-q^i)}{[\![i]\!]^2} + \frac{p}{q}z\left(\frac{2}{[\![i]\!]^2}\sum_{1\leq j\leq i}\frac{q^j}{[\![j]\!]} + \frac{1}{[\![i]\!]}\sum_{1\leq j\leq i}\frac{q^j}{[\![j]\!]^2} - \frac{1}{[\![i]\!]}\sum_{1\leq j\leq i}\frac{q^j}{[\![j]\!]}\right)$$
$$+ pz\left(\frac{2}{[\![i]\!]^2}\sum_{0\leq j<i}\frac{q^j}{[\![j]\!]} + \frac{1}{[\![i]\!]}\sum_{0\leq j<i}\frac{q^j}{[\![j]\!]^2} - \frac{1}{[\![i]\!]}\sum_{0\leq j<i}\frac{q^j}{[\![j]\!]}\right)$$

and

$$d_i(z) = \frac{1+pz}{[\![i]\!]} - \frac{pq^iz}{[\![i]\!]^2} + (Q-q)\frac{z}{[\![i]\!]}\sum_{1\leq j\leq i}\frac{q^j}{[\![j]\!]}.$$

The next step is to plug the values of a_i and d_i in Eq. (11) and express $H(z)$ as a linear combination of "standard" sums $\mathcal{S}_i(z)$, $i = 1, \ldots, 15$, as listed in the appendix. The argument z in each $\mathcal{S}_i(z)$ is omitted for brevity:

$$H(z) = 2(Q-q)[\mathcal{S}_6 + 2pz\mathcal{S}_6 - \mathcal{S}_7 - 3p\mathcal{S}_8 + p\mathcal{S}_9]$$
$$+ 2(Q-q)^2[\mathcal{S}_3 + 2\mathcal{S}_4 - \mathcal{S}_5] + 2\frac{p}{q}(1+pz)^2\mathcal{S}_{10} + 4(Q-q)\frac{p}{q}\mathcal{S}_2$$
$$- 2\frac{p^2}{q^2}(1+pz)\mathcal{S}_{11} + 2\frac{p^2}{q^2}(1+pz)\mathcal{S}_{12} + 2\frac{p^3}{q^3}(1+q)^2\mathcal{S}_1$$
$$+ 4\frac{p^3}{q^2}(1+q)z\mathcal{S}_2 - 2\frac{p^3}{q^3}(1+q)\mathcal{S}_{13} - 2\frac{p^3}{q^2}(1+q)\mathcal{S}_{14} + 2\frac{p^3}{q^2}\mathcal{S}_{15}.$$

Now we give an example of how to obtain the list of coefficients for the sums $\mathcal{S}_i$ from the appendix:

$$[z^n]\,\mathcal{S}_{12}(z) = (-1)^n[w^n](1-w)^{n+2}\sum_{i\geq 1}\frac{w^2q^{2i}}{(1-wq^i)^2(1-wq^iQ)}$$
$$= \sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k[w^{k-2}]\sum_{i\geq 1}\frac{w^2q^{2i}}{(1-wq^i)^2(1-wq^iQ)}$$
$$= \sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}[w^{k-2}]\frac{w^2}{(1-w)^2(1-wQ)}$$
$$= \sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\left(\frac{Q^{k-2}-1}{(Q-1)^2} - \frac{k-2}{Q-1}\right)$$
$$= \frac{1}{(Q-1)^2}\left(-1 + (n+2) - \binom{n+2}{2}\right)$$
$$- \frac{1}{Q-1}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1}.$$

From this, the formula in the appendix is immediate.

Once we have expressed the nth coefficient of each $\mathscr{S}_i$ as an alternating sum, we can compute its asymptotic behavior using Rice's method. The continuation to the complex plane of the discrete sequences that appear in the alternating sums is almost straightforward, except for the sequences of the type

$$\sum_{m=1}^{k} \frac{1}{Q^m - 1}. \tag{12}$$

To get an analytic continuation of this kind of sequence, we write it as a difference of infinite series and shift the index in the second summation:

$$\alpha - \sum_{m \geqslant 1} \frac{1}{Q^{m+k} - 1}, \tag{13}$$

where $\alpha = \sum_{m \geqslant 1} 1/(Q^m - 1)$ is a constant. For instance, for $Q = 2$, the value of α is $1.606695\ldots$

Now, it makes sense to replace k by z in Eq. (13) so the continuation of the sequence to the complex plane is

$$\alpha - \sum_{m \geqslant 1} \frac{1}{Q^{m+z} - 1}.$$

There is a similar sequence of the type $\sum m/(Q^m - 1)$ appearing in the analysis, which could be dealt with in an analogous way, but it turns out that the terms including that kind of sequence cancel out.

The residue computations involved in Rice's method were performed using MAPLE; for the reader's convenience, we will compute the asymptotic behavior of one of the alternating sums containing a sequence of the type given in Eq. (12). There are eight types of alternating sums occurring in $[z^n]\mathscr{S}_1$ up to $[z^n]\mathscr{S}_{15}$. One of these is

$$\sum_{k=3}^{n+2} \binom{n+2}{k} (-1)^k \frac{1}{Q^{k-2} - 1} \sum_{m=1}^{k-3} \frac{1}{Q^m - 1}.$$

We can then write the summation above as a contour integral:

$$-\frac{1}{2\pi i} \int_{\mathscr{C}} \frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)} \frac{1}{Q^{z-2} - 1} \left(\alpha - \sum_{m \geqslant 1} \frac{1}{Q^{m+z-3} - 1} \right) \mathrm{d}z,$$

where $\mathscr{C}$ encloses $\{3, 4, \ldots, n+2\}$. There is a pole of the integrand at $z = 2$ which gives us a main contribution to the asymptotic behavior and there are also poles at $z = 2 + (2\pi i/L)k$, for k an integer different from 0. These last poles contribute a small periodic fluctuation to the coefficient of the leading term in the asymptotic behavior.

We will only consider the pole at $z = 2$ in the discussion that follows. Hence, we have to compute the residue of

$$-\frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)} \frac{1}{Q^{z-2} - 1} \left(\alpha - \sum_{m \geqslant 1} \frac{1}{Q^{m+z-3} - 1} \right).$$

This can be readily done by considering each term separately,

(i) $$-\frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)}\frac{\alpha}{Q^{z-2}-1},$$

(ii) $$\frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)}\frac{1}{(Q^{z-2}-1)^2},$$

(iii) $$\frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)}\frac{1}{Q^{z-2}-1}\frac{1}{Q^{m+z-3}-1},\quad m\geqslant 2,$$

and then sum up the corresponding residues,

(i) $$\frac{\alpha}{2L}n^2\log n+\alpha\left(\frac{\gamma}{2L}-\frac{3}{4L}-\frac{1}{4}\right)n^2+O(n\log n),$$

(ii) $$-\frac{1}{4L^2}n^2\log^2 n+\frac{(2L+3-2\gamma)}{4L^2}n^2\log n$$
$$+\left(\frac{\gamma}{2L}+\frac{3\gamma}{4L^2}-\frac{7}{8L^2}-\frac{3}{4L}-\frac{\pi^2}{24L^2}-\frac{\gamma^2}{4L^2}-\frac{5}{24}\right)n^2+O(n\log^2 n),$$

(iii) $$-\frac{1}{2L}\frac{1}{Q^{m-1}-1}n^2\log n+n^2\left(\frac{1}{Q^{m-1}-1}\left(-\frac{\gamma}{2L}+\frac{3}{4L}+\frac{3}{4}\right)\right.$$
$$\left.+\frac{1}{2(Q^{m-1}-1)^2}\right)+O(n\log n),\quad m\geqslant 2,$$

yielding

$$\operatorname*{Res}_{z=2}\left[-\frac{\Gamma(n+3)\Gamma(-z)}{\Gamma(n+3-z)}\frac{1}{Q^{z-2}-1}\left(\alpha-\sum_{m\geqslant 1}\frac{1}{Q^{m+z-3}-1}\right)\right]$$
$$=-\frac{1}{4L^2}n^2\log^2 n+\frac{(2L+3-2\gamma)}{4L^2}n^2\log n$$
$$+n^2\left(\frac{\alpha+\beta}{2}+\frac{\gamma}{2L}+\frac{3\gamma}{4L^2}-\frac{7}{8L^2}-\frac{3}{4L}-\frac{\pi^2}{24L^2}-\frac{\gamma^2}{4L^2}-\frac{5}{24}\right)+O(n\log^2 n),$$

where α and β are the constants $\sum_{m\geqslant 1}1/(Q^m-1)$ and $\sum_{m\geqslant 1}1/(Q^m-1)^2$, respectively. It turns out that the $n^2\log^2 n$ and $n^2\log n$ terms cancel out when we subtract the square of the expectation to get the variance and only n^2 and lower-order terms remain.

Summing up the asymptotic behavior of each of the alternating sums yields the asymptotic behavior of $E(C_n(C_n-1))$ as $n\to\infty$. Finally, we add $E(C_n)$ and subtract $E^2(C_n)$ in order to get the variance $\mathrm{Var}(C_n)$. For the transparency of the result, we will only give the main ($=n^2$)-term of the result, although in principle one could produce as many lower-order terms as desired.

Theorem 3.1 [Variance of the total unsuccessful search cost (optimized)]. *The variance of the total unsuccessful search cost C_n in a random skip list of n elements as $n \to \infty$, is*

$$\mathrm{Var}(C_n) = n^2(Q-q)^2\left(\frac{\pi^2}{6L^2} + \frac{1}{12} + \frac{1}{L^2} - \frac{1}{2L} - 2(\alpha + \beta)\right)$$

$$+ n^2\left(\frac{Q-q}{QL} + 1 - [\delta_4^2]_0 + \delta_6(\log_Q n)\right) + \mathrm{O}(n \log^2 n),$$

where α and β are the constants $\sum_{m \geqslant 1} 1/(Q^m - 1)$ and $\sum_{m \geqslant 1} 1/(Q^m - 1)^2$, $[\delta_4^2]_0$ is the mean of the square of the periodic function $\delta_4(x)$ (see Theorem 2.1), and $\delta_6(x)$ is another periodic function of period 1 and mean 0. Moreover,

$$\alpha + \beta \approx \frac{\pi^2}{6L^2} - \frac{1}{2L} + \frac{1}{24},$$

for "reasonable" values of q [5].

We now recall the variance of the total successful search cost for the standard search algorithm, for comparison purposes.

Theorem 3.2 [Variance of the total successful search cost (standard)]. *The variance of the total search cost $\tilde{C}_n^{[\mathrm{st}]}$ in a random skip list of n elements, as $n \to \infty$, is*

$$\mathrm{Var}(\tilde{C}_n^{[\mathrm{st}]}) = (Q^2 - 1)n^2\left(\frac{1}{2L} - \frac{\pi^2}{6L^2} + \frac{1}{L^2}\right) + 2(Q-1)n^2\left(\frac{\alpha}{L} - \sum_{m \geqslant 1} \frac{m}{(Q^m - 1)^2}\right)$$

$$+ n^2\left(\frac{\pi^2}{6L^2} + \frac{1}{12} + \varepsilon + \delta_7(\log_Q n)\right) + \mathrm{O}(n \log^2 n),$$

where α is as in the previous theorem, ε is a very small quantity for which a series representation is available, and $\delta_7(x)$ is a periodic function of period 1 and mean 0 [5].

Fig. 5 depicts the coefficients K and K' of n^2 in $\mathrm{Var}(C_n)$ and $\mathrm{Var}(C_n^{[\mathrm{st}]})$, respectively, as a function of q. As we will show in the next section, $\mathrm{Var}(\tilde{C}_n) = Kn^2 + \mathrm{o}(n^2)$ and

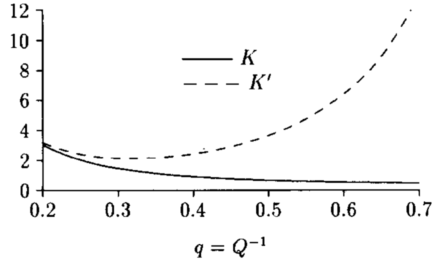

Fig. 5. Behavior of K and K' as a function of q.

Table 1
Some numerical values of K and K'

q	K	K'
0.1	10.79...	10.57...
0.2	3.01...	3.16...
0.3	1.48...	2.15...
0.31...	1.37...	2.13...
0.4	0.93...	2.44...
0.5	0.68...	3.66...
0.6	0.55...	6.41...
0.7	0.48...	12.96...
0.8	0.44...	33.02...
0.9	0.42...	148.13...

hence the comparison of K versus K' makes sense. Table 1 contains several numerical values of both K and K'. The coefficient K' achieves its minimum at $q = 0.31\ldots$

As in the case of the expectation, the coefficient K of the leading term in the asymptotic behavior of the variance does not reach a local minimum, but takes its minimum value for the degenerate case where $q = 1$.

4. Transferring the results to the successful search

As we have already described, C_n and $\tilde{C}_n$ differ only by the number of wRL of the whole sequence $(a_1, \ldots, a_n)$. Our goal is to show that – w.r.t. to the leading terms – the asymptotic behavior of their expectations and variances in the same. We have already shown that this is the case for the expectation of total successful and unsuccessful search cost in Section 2.

We will prove that it is indeed true for the variances from simple properties of probability theory. Assume that X is a random variable with a mean of order $n \log n$, a standard deviation of order n, and maximum of order n^2. And Y is a random variable with a mean of order $\log n$, a standard deviation of order $\log^{1/2} n$, and maximum of order n. Clearly X plays the role of $\tilde{C}_n$, Y that of the number of right-to-left maxima of $(a_1, \ldots, a_n)$ and $X + Y$ that of C_n. Since

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E(XY) - 2E(X)\cdot E(Y),$$

we are done when we can show that $E(XY)$ is of a neglectible order of growth.

We use Chebyshev's inequality in the form

$$\mathrm{Prob}(|Z - E(Z)| > r\sigma) < \frac{1}{r^2}, \quad r > 1.$$

We distinguish cases: First we consider the case where the first parameter is large, $|X - E(X)| > n^{\alpha+1}$. This has a small probability, and we bound both parameters by their maxima. In the other case, we distinguish the two subcases, where the second parameter is large, $|Y - E(Y)| > n^{\beta} \log^{1/2} n$, and we bound the second parameter by its maximum, and the remaining subcases, where we cannot say anything about the probabilities, but both parameters are small.

Doing as indicated, we find

$$E(XY) = O(\text{Prob}(|X - E(X)| > n^{\alpha+1})) \cdot O(n^2) \cdot O(n)$$
$$+ O(n^{1+\alpha}) \cdot O(|\text{Prob}(|Y - E(Y)| > n^{\beta} \log^{1/2} n)) \cdot O(n)$$
$$+ O(n^{1+\alpha}) \cdot O(n^{\beta} \log^{1/2} n) \cdot O(n).$$

By Chebyshev's inequalities we find

$$E(XY) = O(n^{3-2\alpha}) + O(n^{2+\alpha-2\beta}) + O(n^{1+\alpha+\beta+\varepsilon}),$$

where we replaced the logarithmic factor by n^{ε}, for simplicity.

Now we can obtain a relatively small error term by balancing the three exponents. It turns out that $\alpha = \frac{5}{9}$ and $\beta = \frac{1}{3}$ is the optimal choice, and then our bound is $O(n^{17/9+\varepsilon})$.

This is probably far from the expected sharpest estimate, since the next term in the asymptotic expansion of $\text{Var}(\tilde{C}_n)$ should also be of order $n \log^2 n$. However, we do not see any other way of proving this than to do most of the computations also in this instance, and since the recursion is not so nice, they would be even messier, whence we decided to confine ourselves with this elementary bounding technique.[3]

5. Extensions

An important property of the total unsuccessful search cost analyzed in previous sections is that it is an *additive cost*. We are specially interested in the methodological aspects of the analysis of such costs.

Definition 5.1. A cost C over skip lists is said to be an *additive cost* if, given a skip list S of height m, C satisfies

$$C(S) = C(\sigma) + C(\tau) + U(|\sigma|, |\tau|, m),$$

where σ is the part of S before the partitioning element, and τ is the part of S after the partitioning element. The function U is called a *valuation* function over skip lists.

Note that this definition as well as the forecoming developments apply equally well for sequences of i.i.d. random variables.

[3] In the meantime Kirschenhofer has shown that in fact $\text{Var}(C_n) - \text{Var}(\tilde{C}_n) = O(n \log n)$.

The main interest of additive costs stems from the fact that the algebraic part of the probabilistic analysis can be carried in a uniform way. Moreover, it can be carried out even if the levels a_j of the elements of skip lists were generated by independent random variables other than geometric ones. In order to do that, let us denote $\pi_k = \text{Prob}(a_j = k)$ and $\phi_k = \text{Prob}(a_j \leqslant k)$. Furthermore, let

$$[\![m]\!] := 1 - z\phi_m.$$

Note that $[\![m]\!] = [\![m-1]\!] - \pi_m z$ and that this definition is consistent with the one that we give in Section 2 for the case of geometrically distributed levels.

We first introduce a family of probability generating functions $P^*(z, y)$:

$$P^*(z,y) = \sum_{s \in \mathscr{S}^*} \text{Prob}(s)\, y^{C(s)} z^{|s|},$$

where $\mathscr{S}^*$ denotes the set of skip lists whose height satisfies condition (*).

If we use the recursive decomposition skip lists and since C is additive, we have,

$$\begin{aligned} P^{=m}(z,y) &= \sum_{\substack{\sigma \in \mathscr{S}^{<m} \\ \tau \in \mathscr{S}^{\leqslant m}}} \text{Prob}(\sigma)\,\text{Prob}(\tau)\,\pi_m\, y^{C(\sigma)+C(\tau)+U(|\sigma|,|\tau|,m)} z^{|\sigma|+|\tau|+1} \\ &= \pi_m z \Bigg(\sum_{\substack{\sigma \in \mathscr{S}^{<m} \\ \tau \in \mathscr{S}^{\leqslant m}}} \text{Prob}(\sigma)\,\text{Prob}(\tau)\, y^{U(|\sigma|,|\tau|,m)} y^{C(\sigma)} y^{C(\tau)} z^{|\sigma|} z^{|\tau|} \Bigg). \end{aligned}$$

If we differentiate with respect to y and set $y = 1$ we have

$$\begin{aligned} S^{=m}(z) &= \left.\frac{\partial P^{=m}}{\partial y}\right|_{y=1} \\ &= \pi_m z \Bigg(S^{<m}(z) \sum_{\tau \in \mathscr{S}^{\leqslant m}} \text{Prob}(\tau) z^{|\tau|} + S^{\leqslant m}(z) \sum_{\sigma \in \mathscr{S}^{<m}} \text{Prob}(\sigma) z^{|\sigma|} + U_m(z) \Bigg) \\ &= \pi_m z \left(\frac{S^{<m}(z)}{[\![m]\!]} + \frac{S^{\leqslant m}(z)}{[\![m-1]\!]} + U_m(z) \right), \end{aligned}$$

where $U_m(z)$ is defined as follows:

$$U_m(z) = \sum_{\substack{\sigma \in \mathscr{S}^{<m} \\ \tau \in \mathscr{S}^{\leqslant m}}} \text{Prob}(\sigma)\,\text{Prob}(\tau)\, U(|\sigma|,|\tau|,m)\, z^{|\sigma|+|\tau|}.$$

If we assume w.l.o.g. that $S^{=0}(z) = 0$ then we can solve the linear recurrence using iteration and get

$$S^{\leqslant m}(z) = \frac{z}{[\![m]\!]^2} \sum_{1 \leqslant i \leqslant m} [\![i]\!]\,[\![i-1]\!]\,\pi_i U_i(z).$$

The quantity we are interested in, namely the expectation of C over skip lists of size n, is the nth coefficient of the limit generating function $S(z)$:

$$S(z) = \lim_{m\to\infty} C^{\leqslant m}(z) = \frac{z}{(1-z)^2} \sum_{i\geqslant 1} [\![i]\!] [\![i-1]\!] \pi_i U_i(z). \tag{14}$$

A particularly easy and nice situation arises when the valuation function U is of the type $U(|\sigma|,|\tau|,m) = f(|\sigma|) + g(|\tau|)$, and both f and g have Taylor series expansions around $x = 0$, namely

$$f(x) = f_0 + f_1 x + f_2 x^2 + \cdots, \qquad g(x) = g_0 + g_1 x + g_2 x^2 + \cdots$$

Then

$$S(z) = \frac{z}{(1-z)^2} \sum_{i\geqslant 1} \pi_i \left([\![i-1]\!] f(\vartheta)\left(\frac{1}{[\![i-1]\!]}\right) + [\![i]\!] g(\vartheta)\left(\frac{1}{[\![i]\!]}\right)\right), \tag{15}$$

where $\vartheta := z(\mathrm{d}/\mathrm{d}z)$, I is the identity operator and $f(\vartheta)$ and $g(\vartheta)$ are the operators

$$f(\vartheta) = f_0 I + f_1 \vartheta + f_2 \vartheta^2 + \cdots, \qquad g(\vartheta) = g_0 I + g_1 \vartheta + g_2 \vartheta^2 + \cdots$$

Moreover, for $T(z) = \lim_{m\to\infty} P_{yy}^{\leqslant m}(z,1)$, the generating function for the second factorial moments, we have

$$\begin{aligned} T(z) = \frac{z}{(1-z)^2} \sum_{i\geqslant 1} \pi_i \Bigg[& [\![i-1]\!] \left(2f(\vartheta)(S^{<i}(z)) + (f^2 - f)(\vartheta)\left(\frac{1}{[\![i-1]\!]}\right)\right) \\ & + [\![i]\!] \left(2g(\vartheta)(S^{\leqslant i}(z)) + (g^2 - g)(\vartheta)\left(\frac{1}{[\![i]\!]}\right)\right) \\ & + 2\left(S^{<i}(z) + f(\vartheta)\left(\frac{1}{[\![i-1]\!]}\right)\right)\left(S^{\leqslant i}(z) + g(\vartheta)\left(\frac{1}{[\![i]\!]}\right)\right)\Bigg]. \end{aligned} \tag{16}$$

Similar techniques work for additive costs that are defined in terms of a reverse recursive decomposition of the skip lists: skip lists of height m are decomposed into two skip lists of heights $\leqslant m$ and $< m$, respectively, using the last element of level m of the skip list as the partitioning element.

Furthermore, we can also introduce and apply the former algebraic techniques for costs satisfying linear recurrences, i.e.

$$C(s) = \alpha C(\sigma) + \beta C(\tau) + U(|\sigma|,|\tau|).$$

A particular case of these linear additive costs are the costs where either α or β is 0. Important examples of such type of costs are the number of left-to-right and right-to-left maxima in a skip list S.

Appendix

Here is the list of sums that we used in our analysis. They are not sorted in any particular order:

$$\mathscr{S}_1(z)=\frac{z^3}{(1-z)^2}\sum_{\substack{1\leqslant j\leqslant i\\ 1\leqslant h<i}}\frac{q^{i+j+h}}{[\![i]\!][\![i-1]\!][\![j]\!][\![h]\!]},\qquad \mathscr{S}_9(z)=\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{2i}}{[\![i]\!]^2},$$

$$\mathscr{S}_2(z)=\frac{z^2}{(1-z)^2}\sum_{1\leqslant j<i}\frac{q^{i+j}}{[\![i]\!][\![i-1]\!][\![j]\!]},\qquad \mathscr{S}_{10}(z)=\frac{z}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{i}}{[\![i]\!][\![i-1]\!]},$$

$$\mathscr{S}_3(z)=\frac{z^2}{(1-z)^2}\sum_{1\leqslant j\leqslant i}\frac{q^{i+j}}{[\![i]\!][\![j]\!]^2},\qquad \mathscr{S}_{11}(z)=\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{2i}}{[\![i]\!][\![i-1]\!]^2},$$

$$\mathscr{S}_4(z)=\frac{z^2}{(1-z)^2}\sum_{1\leqslant j\leqslant i}\frac{q^{i+j}}{[\![i]\!]^2[\![j]\!]},\qquad \mathscr{S}_{12}(z)=\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{2i}}{[\![i]\!]^2[\![i-1]\!]},$$

$$\mathscr{S}_5(z)=\frac{z^2}{(1-z)^2}\sum_{1\leqslant j\leqslant i}\frac{q^{i+j}}{[\![i]\!][\![j]\!]},\qquad \mathscr{S}_{13}(z)=\frac{z^3}{(1-z)^2}\sum_{1\leqslant j\leqslant i}\frac{q^{2i+j}}{[\![i]\!][\![i-1]\!]^2[\![j]\!]},$$

$$\mathscr{S}_6(z)=\frac{z}{(1-z)^2}\sum_{1\geqslant 1}\frac{q^{i}}{[\![i]\!]^2},\qquad \mathscr{S}_{14}(z)=\frac{z^3}{(1-z)^2}\sum_{1\leqslant j<i}\frac{q^{2i+j}}{[\![i]\!]^2[\![i-1]\!][\![j]\!]},$$

$$\mathscr{S}_7(z)=\frac{z}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{i}}{[\![i]\!]},\qquad \mathscr{S}_{15}(z)=\frac{z^3}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{3i}}{[\![i]\!]^2[\![i-1]\!]^2},$$

$$\mathscr{S}_8(z)=\frac{z^2}{(1-z)^2}\sum_{i\geqslant 1}\frac{q^{2i}}{[\![i]\!]^3}.$$

Next, we give the list of the coefficients $[z^n]\,\mathscr{S}_1,\dots,[z^n]\,\mathscr{S}_{15}$:

$$[z^n]\,\mathscr{S}_1(z)=-\frac{1}{Q-1}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\left[\binom{k-3}{2}+\frac{k-3}{Q-1}+\sum_{m=1}^{k-3}\frac{m-2}{Q^m-1}\right],$$

$$[z^n]\,\mathscr{S}_2(z)=\frac{1}{Q-1}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-3}{Q^{k-2}-1},$$

$$[z^n]\,\mathscr{S}_3(z)=\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\left[\binom{k-2}{2}-\sum_{m=1}^{k-3}\frac{m}{Q^m-1}+(k-2)\sum_{m=1}^{k-3}\frac{1}{Q^m-1}\right],$$

$$[z^n]\mathscr{S}_4(z) = \sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\left[\binom{k-2}{2}+\sum_{m=1}^{k-3}\frac{m}{Q^m-1}\right],$$

$$[z^n]\mathscr{S}_5(z) = -\sum_{k=2}^{n+1}\binom{n+1}{k}(-1)^k\frac{1}{Q^{k-1}-1}\left[k-2+\sum_{m=1}^{k-2}\frac{1}{Q^m-1}\right],$$

$$[z^n]\mathscr{S}_6(z) = -\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1},$$

$$[z^n]\mathscr{S}_7(z) = \sum_{k=2}^{n+1}\binom{n+1}{k}(-1)^k\frac{1}{Q^{k-1}-1},$$

$$[z^n]\mathscr{S}_8(z) = \sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\binom{k-2}{2},$$

$$[z^n]\mathscr{S}_9(z) = -\sum_{k=2}^{n+1}\binom{n+1}{k}(-1)^k\frac{k-2}{Q^{k-1}-1},$$

$$[z^n]\mathscr{S}_{10}(z) = \frac{1}{Q-1}\binom{n+1}{2},$$

$$[z^n]\mathscr{S}_{11}(z) = \frac{1}{(Q-1)^2}\binom{n+1}{2}-\frac{n}{Q(Q-1)}$$
$$+\frac{1}{Q(Q-1)}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1},$$

$$[z^n]\mathscr{S}_{12}(z) = -\frac{1}{(Q-1)^2}\binom{n+1}{2}-\frac{1}{Q-1}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1},$$

$$[z^n]\mathscr{S}_{13}(z) = \frac{n^2}{(Q-1)^3}-\frac{n}{Q(Q-1)^3}+\frac{1}{(Q-1)^2}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-3}{Q^{k-2}-1}$$
$$-\frac{Q-q}{(Q-1)^3}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1}$$
$$-\frac{1}{Q(Q-1)}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\left[\binom{k-2}{2}\right.$$
$$\left.+\sum_{m=1}^{k-3}\frac{m}{Q^m-1}\right],$$

$$[z^n]\mathscr{S}_{14}(z) = \frac{1}{Q-1}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{1}{Q^{k-2}-1}\sum_{m=1}^{k-3}\frac{m}{Q^m-1}$$
$$-\frac{1}{(Q-1)^2}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-3}{Q^{k-2}-1}$$

$$[z^n]\mathscr{S}_{15}(z) = -\frac{n^2}{(Q-1)^3} - \frac{n}{Q(Q-1)^3} - \frac{Q-q}{(Q-1)^3}\sum_{k=3}^{n+2}\binom{n+2}{k}(-1)^k\frac{k-2}{Q^{k-2}-1}.$$

References

[1] L. Devroye, A limit theory for random skip lists, *The Ann. Appl. Probab.* **2** (1992) 597–609.

[2] Ph. Flajolet and R. Sedgewick, Mellin transforms and asymptotics: finite differences and Rice's integrals, *Theoret. Comput. Sci.* **144** (1995) 101–124, this volume.

[3] J. Gabarró, C. Martínez and X. Messeguer, A design of a parallel dictionary using skip lists, *Theoret. Comput. Sci.* (1996), to appear.

[4] I. Goulden and D. Jackson, *Combinatorial Enumeration* (Wiley, New York, 1983).

[5] P. Kirschenhofer and H. Prodinger, The path length of random skip lists, *Acta Inform.* (1994), to appear.

[6] T. Papadakis, Skip lists and probabilistic analysis of algorithms, Ph.D. Thesis, University of Waterloo, 1993; available as Tech. Report CS-93-28.

[7] T. Papadakis, J.I. Munro and P.V. Poblete, Average search and update costs in skip lists, *BIT* **32** (1992) 316–332.

[8] H. Prodinger, Combinatorics of geometrically distributed random variables: left-to-right maxima, *Discrete Math.*, to appear.

[9] W. Pugh, A skip list cookbook, Tech. Report CS-TR-2286.1, Institute for Advanced Computer Studies, Department of Computer Science, University of Maryland, College Park, MD, 1990; also published as UMIACS-TR-89-72.1.

[10] W. Pugh, Skip lists: a probabilistic alternative to balanced trees, *Comm. ACM*, **33** (1990) 668–676.

[11] U. Schmid, On a tree collision resolution algorithm in presence of capture, *RAIRO Informatique Théorique et Applications/Theoret. Inform. Appl.* **2** (1992) 163–197.