ssdnm
środowiskowe
studia doktoranckie
z nauk matematycznych

Michał Dębski

Uniwersytet Warszawski

Optimal k-radius sequences

Praca semestralna nr 1

(semestr letni 2011/12)

Opiekun pracy: Jarosław Grytczuk

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

OPTIMAL k-RADIUS SEQUENCES

MICHAŁ DĘBSKI

Abstract. A k-radius sequence over an alphabet A is a sequence in which every pair of elements from A occurs at least once within distance k. By $f_k(n)$ we denote the length of a shortest k-radius sequence over an n-ary alphabet. This problem was originaly studied by Jaromczyk and Lonc, and emerged from the problem of evaluating a two argument function on all pairs of n large objects [3].

In this paper we survey some results concerning the asymptotic behaviour of $f_k(n)$. We base on results by Jaromczyk and Lonc [3] (basic properties of k-radius sequences), Blackburn [1] (existential upper bound on $f_k(n)$ for constant k) and Jaromczyk, Lonc and Truszczynski [4] (constructive upper bounds in cases when k is constant and equal to n^x, where $0 < x < 1$, and optimal construction for $k = 2$).

In particular, we show that in both cases $f_k(n) \sim \frac{n^2}{2k}$ and bound the nonsignicifant terms by $O(n^{1+\epsilon})$ when k is fixed and by $O(n^\beta)$ for some $\beta < 2 - x$ when $k = n^x$. In the latter case we obtain a constant β that is slightly smaller than the one obtained by Jaromczyk, Lonc and Truszczyñski [4] due to some technical improvements.

1. Introduction

The study of k-radius sequences originated from the problem of comparing a set of large objects [3]. Suppose that we need to compute a two-argument function on all pairs of n objects $(o_1, \ldots, o_n)$, that are accessible by *read* operation, where a single read places one of the objects in memory. Assuming that the memory can store up to $k + 1$ objects at the same time, what is the minimum number of reads that we need to perform?

We will consider only the FIFO (First In First Out) memory management. That is, when the memory is full, before the read operation we remove an "oldest" object, obtained k+1 reads earlier. In this case we can reformulate our problem as finding a shortest possible sequence of read operations such that for any pair (i, j) at most $k - 1$ other objects are read between some two reads of o_i and o_j.

As an example, suppose that we have 6 objects (denoted as numbers from 1 to 6) and our memory can store only 3 of them. Consider a sequence of reads $s = 123456124536$. Any two elements can be found in some window of length 3 (or, equivalently, within distance 2), so s is a 2-*radius sequence*. As we will see later (Lemma 4), it is the shortest such sequence.

In section 2 we give a formal definition of the problem and establish some simple preliminary results, following [3] and [4]. Section 3 is devoted to bounds on the length of optimal k-radius sequence - we give simple lower bounds (from [3]) and a nonconstructive upper bound, that uses a result on covering of hypergraphs [1].

We will be most interested in constructions of optimal, or almost optimal, k-radius sequences. In section 4 we give optimal constructions of 1-radius sequences

over an arbitrary alphabet and 2-radius sequences over an alphabet of size $2p$ (where p is prime) [4].

The main results are stated in sections 5, 6 and 7. In section 5 we show an explicit construction that leads to a recursive upper bound on an optimal sequences (Corollary 13). In the following two sections we establish explicit upper bounds for the cases when k is constant and equal n^α (where $0 < \alpha < 1$). We follow the argument from [4], but our result is slightly stronger due to some technical improvements.

Finally, in section 8 we briefly discuss some open problems concenrning k-radius sequences and relate this topic to some other areas of interest.

2. Definitions and basic properties

Let $s = s_1 s_2 \dots s_m$ be a sequence of elements from the set A of size n. We say that s is a k-radius sequence over A if every two elements of A occur in s within distance of at most k, that is, for any $a, b \in A$ there exist indices i, j such that $s_i = a$, $s_j = b$ and $|i - j| \leq k$. We denote the length of the shortest possible k-radius sequence over A by $f_k(n)$, where k can be either a constant or a function of n.

Note that a k-radius sequence is also a $(k+1)$-radius sequence. Indeed, if all pairs from A occur within distance at most k, they occur within distance at most $k+1$. Therefore, we have

Lemma 1. *For any n and k, k' such that $k' < k$ we have*

$$f_k(n) \leq f_{k'}(n).$$

□

On the other hand, we can remove from a given k-radius sequence the most common symbols, obtaining a k-radius sequence over a smaller alphabet:

Lemma 2. *For any k and n, n' such that $n' < n$ we have*

$$f_k(n') \leq \frac{n'}{n} f_k(n).$$

Proof. Take s to be a k-radius sequence of length $f_k(n)$ over A. The $n - n'$ most common symbols occur in s a total of at least $\left\lceil \frac{n-n'}{n} f_k(n) \right\rceil$ times. By deleting those symbols we obtain a sequence of length $f_k(n') - \left\lceil \frac{n-n'}{n} f_k(n) \right\rceil = \left\lfloor \frac{n'}{n} f_k(n) \right\rfloor$. □

A useful observation is that if we take a k-radius sequence over A and replace each element $a \in A$ with a sequence of lenght (at most) m, we obtain a $((k+1)m - 1)$-radius sequence over an alphabet of size $m\,|A|$.

Lemma 3. *For any k, n and $m > 0$ we have*

$$f_{(k+1)m}(mn) \leq m f_k(n).$$

Proof. Take s to be a k-radius sequence of length $f_k(n)$ over $A = \{a_1, a_2, \dots, a_n\}$. Let $A_1, A_2, \dots, A_n$ be n disjoint sets of size m. Take s' to be a sequence obtained from s by replacing each element a_i by a permutation of the set A_i. Clearly, the length of s' is $m f_k(n)$.

Note that any two elements of A_i occur in s' within distance at most $m - 1 \leq (k+1)m$. Also, for any i, j we have occurrences of a_i and a_j in s separated by at

most $k-1$ other symbols. Corresponding permutations of A_i and A_j are separated by at most $(k-1)m$ symbols in s', so any pair from $A_i \times A_j$ is separated by at most $(k-1)m + |A_i| + |A_j| - 2 = (k+1)m - 2$ other symbols. Therefore s' is a $(k+1)m$-radius sequence. □

3. General bounds for $f_k(n)$

A lower bound for $f_k(n)$ is obtained by counting how many times each symbol must appear in an optimal sequence:

Lemma 4. *For any n, k we have*

$$f_k(n) \geq \left\lceil \frac{n-1}{2k} \right\rceil n.$$

Proof. Note that for any position i in a sequence there are at most $2k$ other positions within distance k from i, so an element at position i can see at most $2k$ other elements. In order for an element a to see other $n-1$ elements, it must be placed in at least $\lceil \frac{n-1}{2k} \rceil$ positions in a sequence. As there are n elements, the result follows. □

Alternatively, we may count the number m of pairs of positions in a sequence of length l that are within distance k. The value of m must be at least $\binom{n}{2}$ in order to cover all pairs from A. As m is equal to $lk - \frac{k(k+1)}{2}$, we obtain the following lemma.

Lemma 5. *For any n, k we have*

$$f_k(n) \geq \frac{1}{k}\binom{n}{2} + \frac{k+1}{2}.$$

□

For the sake of analyzing the asymptotic behaviour of $f_k(n)$ we rewrite this result as follows.

Corollary 6. *For any k we have*

$$f_k(n) \geq \frac{n^2}{2k} + O(n).$$

□

Before proving the upper bound on $f_k(n)$ we need to formulate a result on covering of hypergraphs. A cover of a hypergraph H is a set C of hyperedges of H such that every vertex is contained in at least one hyperedge from C. The degree of a vertex v, denoted $deg(v)$, is the number of edges of H that contain v and similarily the codegree of vertices u, v, denoted $codeg(u, v)$, is the number of edges that contain both u and v.

Note that the optimal (minimum) covering of a r-uniform hypergraph on n vertices must contain at least $\frac{n}{r}$ hyperedges. The result that we use basically says that we can get arbitrarily close to that number, provided that our hypergraph is large enough, all vertices have the same degree d, and codegrees are small compared to d.

Theorem 7 ([1, Theorem 3])**.** *Fix an integer r and a positive real number δ. Then there are n_0 and δ' such that:*

If H is an r-uniform hypergraph on $n > n_0$ vertices, such that every vertex have degree d and $codeg(u,v) \leq \delta' d$ for any pair of distinct vertices u, v, then there exists a cover of H consisting of at most $(1+\delta)\frac{n}{r}$ edges.

We will use this result to find a covering of a hypergraph H, defined as follows. The vertices of H are all (onordered) pairs of distinct elements from A. Edges of H correspond to all sequences t of length l, that consist of l distinct elements from A, where an edge e_t contains all pairs that are within distance k in a sequence t. An almost optimal covering of H defines an almost optimal k-radius sequence - a concatenation of sequences corresponding to selected hyperedges.

Theorem 8 ([1, Theorem 1])**.** *For any constant k we have*

$$f_k(n) \leq \frac{1}{k}\binom{n}{2} + o(n^2).$$

Proof. In order to prove the theorem, we need to show that for any $\epsilon > 0$ there exists n_0 such that for any $n > n_0$ there is a k-radius sequence over n-ary alphabet of length at most $(1+\epsilon)\frac{1}{k}\binom{n}{2}$.

For any l and $n > l$ we define the hypergraph H_n^l on a set of $\binom{n}{2}$ (unordered) pairs of distinct elements from some set A of cardinality n. Hyperedges of H_n^l correspond to all sequences over A of length l with all entries distinct, where a hyperedge e_t corresponding to a sequence $t = t_1 t_2 \dots t_l$ contains all pairs that are within distance k in t (that is, $e_t = \bigcup_{i=1}^{l-k}\bigcup_{j=1}^{k}\{t_i, t_{i+j}\}$).

Note that H_n^l is r-uniform, where r is the number of pairs of positions that are within distance k in a sequence of length l, and $r = lk - \frac{k(k+1)}{2}$ does not depend on n. Note that in our construction the number of sequences that cover a pair $\{a,b\} \subset A$ is the same for all a and b, so all vertices of H_n^l have the same degree d.

Similarly, fix $a, b \in A$. Clearly, for any $c \in A \setminus \{a,b\}$ the value $codeg(\{a,b\},\{a,c\})$ is the same. Summing over all possible c and over all edges that cover $\{a,b\}$ we get that $(n-2)codeg(\{a,b\},\{a,c\}) = dr$, so the codegree of two interseting pairs equals $\frac{dr}{n-2}$. The same reasoning shows that codegree of two non intersecting pairs is $\frac{dr}{\binom{n-2}{2}} \leq \frac{dr}{n-2}$.

Now, fix $\epsilon > 0$ and choose l such that $l \leq \frac{r}{k}\sqrt{1+\epsilon}$. Let n_0 and δ' be the values from Theorem 7 for the parameter $\delta = \sqrt{1+\epsilon} - 1$. Take $n_0' = \max(n_0, \frac{r}{\delta'} + 2, l+1)$. Take $n \geq n_0'$, and recall that H_n^l is an r-uniform hypergraph on $\binom{n}{2} \geq n_0$ vertices, such that every vertex have degree d and $codeg(u,v) \leq \frac{dr}{n-2} \leq \delta' d$, for any pair of distinct vertices u, v. By Theorem 7, there exists a cover $C = \{e_1, e_2, \dots, e_s\}$ of H_n^l such that $s \leq \sqrt{1+\epsilon}\frac{\binom{n}{2}}{r}$.

Let $t_1, t_2, \dots, t_s$ be the sequences corresponding to the edges $e_1, e_2, \dots, e_s$. As C is a cover of H_n^l of order s, the concatenation $t_1 t_2 \dots t_s$ is a k-radius sequence over A of length

$$ls \leq (\frac{r}{k}\sqrt{1+\epsilon})(\sqrt{1+\epsilon}\frac{\binom{n}{2}}{r}) = \frac{1}{k}\binom{n}{2}(1+\epsilon)$$

which completes the proof.

□

Note that the leading terms in Corollary 6 and Theorem 8 are the same, so for constant k our bounds are asymptotically tight:

Corollary 9. *For any constant k we have*

$$f_k(n) = \frac{n^2}{2k} + o(n^2).$$

4. Construction of optimal sequences for small k

In some cases we can construct k-radius sequeces that achieve the lower bound from Lemma 4 or 5. In this section we show two constructions when $k = 1$ for any n, and $k = 2$ for $n = 2p$ (where p is prime).

When $k = 1$, we may think of a 1-radius sequence as a walk in K_n that covers all edges. We can construct such a sequence using Eulerian cycle in K_n for odd n.

Theorem 10 ([3, Theorem 1])**.**

$$f_1(n) = \begin{cases} \binom{n}{2} + 1, & \text{for } n \text{ odd,} \\ \binom{n}{2} + \frac{1}{2}n, & \text{for } n \text{ even.} \end{cases}$$

Proof. Take n to be odd. Consider a complete graph G on a set of vertices $A = \{a_1, a_2, \ldots, a_n\}$. As all degrees in G are even, there exists an Euler cycle $C = a_{i_1}a_{i_2}\ldots a_{i_m}$, where $m = \binom{n}{2}$. Note that in C (thought of as a sequence) every pair of distinct elements from A occur within distance 1, except for the pair a_{i_1}, a_{i_m}. Therefore, a sequence $Ca_{i_1} = a_{i_1}a_{i_2}\ldots a_{i_m}a_{i_1}$ is a 1-radius sequence of length $\binom{n}{2} + 1$, which is optimal by Lemma 5.

Now, we will bouild a 1-radius sequence over the set $A \cup \{x\}$ (of size $n + 1$). Take $R = a_1xa_2a_3xa_4\ldots a_{n-1}a_nx$. Note that R covers all pairs of the form $\{x, a_i\}$. Without loss of generality me may assume that the sequence C (defined above) starts with a_1. Therefore, the sequence CR is a 1-radius sequence over an alphabet of size $n + 1$. As R is of length $3\frac{n-1}{2} + 2$, we constructed a 1-radius sequence of length $3\frac{n-1}{2} + 2 + \frac{n(n-1)}{2} = \frac{n^2+2n+1}{2} = (n+1)\frac{(n+1)}{2} = (n+1)\lceil\frac{n}{2}\rceil$, which is optimal by Lemma 4. Note that $\binom{n+1}{2} + \frac{1}{2}(n+1) = \frac{n^2+2n+1}{2}$, which completes the proof. □

In our construction for $k = 2$ and $n = 2p$ (where p is an odd prime) we represent an alphabet A as a product of additive groups $\mathbb{Z}_2 \times \mathbb{Z}_p$. We construct a number of (hamiltonian) cycles H_j that consist of edges of the form $(a, b)(a + 1, b + j)$, for $j = 1, 2, \ldots, \frac{p-1}{2}$ (that cover pairs of the form $\{(a, b), (a+1, b\pm j)\}$ and $\{(a, b), (a, b\pm 2j)\}$). A 2-radius sequence is constructed by combining fragments of those cycles and adding a sequence that covers the remaining pairs (that are no longer covered by cycle fragments and those of the form $\{(a, b), (a + 1, b)\}$).

Theorem 11 ([4, Theorem 5.6 and Corollary 5.7])**.** *For any odd prime p we have*

$$f_2(2p) = p^2 + p.$$

Proof. Take an alphabet $A = \mathbb{Z}_2 \times \mathbb{Z}_p$. For any $j \in \{1, 2, \ldots, \frac{p-1}{2}\}$ consider a (cyclic) sequence H_j defined by $H_j(i + 1) = H_j(i) + (1, j)$ (where the addition is performed modulo 2 on first coordinate and modulo p on the second) and $H_j(0) = (0, 0)$. Note that $H_j(i) = H_j(i + m)$ requires that m is divisible by 2 and by p, as $H_j(i + m) - H_j(i) = (m(\mathrm{mod}\ 2), jm(\mathrm{mod}\ p))$ and j is prime to p. Therefore, H_j runs through all $2p$ elements of A.

Any pair of elements from A not of the form $\{(0,b),(1,b)\}$ occurs in some H_j within distance 2. Indeed, for $(a,b),(a+1,b')$, either $r=b-b'$, or $r=b'-b$ (modulo p) is nonzero and not larger than $\frac{p-1}{2}$, so those elements are consecutive in H_r. Similarily, for $(a,b),(a,b')$ we can find $r\leq\frac{p-1}{2}$ such that $2r=b-b'$ or $2r=b'-b$ and the specified pair occurs in H_r at distance 2.

Now, define S_j to be the sequence that consists of first $2p$ positions in H_j. Note that S_j starts with $(0,0)$, and if we append $(0,0)$ to S_j, the resulting sequence covers the same pairs as H_j, except $\{(1,-j),(1,j)\}$. Therefore, the sequence $S(0,0)$, where $S=S_1S_2\dots S_{\frac{p-1}{2}}$, covers of pairs from A except those of the form $\{(0,b),(1,b)\}$ and $\{(1,j),(1,-j)\}$.

Now, take R to be a concatenation of $(0,0)(0,1)$ and $\frac{p-1}{2}$ blocks of the form $(0,j)(1,j)(1,-j)(0,-j)$. As R covers all remaining pairs and starts with $(0,0)$, the sequence SR has the k-radius property.

The length of constructed sequence equals $2p\frac{p-1}{2}+4\frac{p-1}{2}+2=p^2+p$. As $2p=2(\mathrm{mod}\ 4)$, it coincides with the lower bound $\lceil\frac{2p-1}{4}\rceil 2p=\frac{2p+2}{4}2p$ from Lemma 4, and the proof is finished. □

5. Main construction

In this section we describe the general construction of k-radius sequences, that will be used later to obtain precise asymptotics for $f_k(n)$. After we etablish Lemma 12, the rest of the results will follow by adjusting the parameters.

The construction proceeds as follows: we take as an alphabet the set $\mathbb{Z}_{\not\equiv\neg+\not\equiv}\times\mathbb{Z}_q$ for some (not nessasirily prime) q and construct a number of cyclic sequences c_j^d defined by $c_j^d(i+1)=c_j^d(i)+(1,j)$, for $j=0,1,\dots,q-1$. We argue that the sequences c_j^d cover all pairs except $\{(a,b),(a,b')\}$, provided that q have no small divisors. Next, we make sequences s_j^d that cover the same pairs as c_j^d (by repeating the first k terms) and take the concatenation of all those sequences. Finally, we append $2k+1$ "smaller" k-radius sequences covering the remaining pairs.

Lemma 12 ([4, proof of Lemma 2.7])**.** *For all k and $n=(2k+1)q$, where all divisors of q are greater than k, we have*

$$f_k(n)\leq(2k+1)f_k(\frac{n}{2k+1})+\frac{n^2}{2k+1}+k\sum_{j=0}^{q-1}\gcd(q,(2k+1)j).$$

Proof. Take an alphabet $A=\mathbb{Z}_{2k+1}\times\mathbb{Z}_q$ and consider a sequence c_j^0 defined by $c_j^0(i+1)=c_j^0(i)+(1,j)$ and $c_j^0(0)=(0,0)$, for some $j\in\{0,1,\dots,q-1\}$. Note that $c_j^0(a+i(2k+1))=(a,aj+i(2k+1)j)$, and $i(2k+1)j$ takes $\frac{q}{\gcd(q,(2k+1)j)}$ values modulo q $(d,2d,\dots,(\frac{q}{d}-1)d$, where $d=\gcd(q,(2k+1)j))$. It follows that c_j^0 contains exactly $(2k+1)\frac{q}{d}$ distinct elements.

Now, define (cyclic) sequences c_j^r such that $c_j^r(i)=c_j^0(i)+(0,r)$, for $r=0,1,\dots,d-1$. By the previous argument, c_j^r contains exactly $(2k+1)\frac{q}{d}$ distinct elements. Moreover, for fixed j each element of $\mathbb{Z}_{2k+1}\times\mathbb{Z}_q$ occurs in exactly one of the sequences c_j^r (as $c_j^r(a+i(2k+1))=(a,aj+r+i(2k+1)j)$ and for all a each sequence c_j^r contains elements $\{(a,id+aj+r):i=0,1,\dots,\frac{q}{d}-1\}$).

We claim that sequences c_j^r together cover all pairs from A that are not of the form $\{(a,b),(a,b')\}$. Indeed, take any pair $(a,b),(a',b')$ (where $a'\neq b'$). Without loss of generality we may assume that $a'=a+i$, where $i\in\{1,2,\dots,k\}$. We need

to find j such that $b + ij = b'(\mathrm{mod}\ q)$. By our assumptions i is prime to q, so there is a number i^{-1} such that $ii^{-1} = 1(\mathrm{mod}\ q)$, and we may take $j = (b' - b)i^{-1}$. By our previous argument the pair (a, b) occurs in sequence c_j^r for some r. As $(a', b') = (a, b) + i(1, j)$, the pair (a', b') occurs at distance i from (a, b) which completes the proof of the claim.

Define s_j^r to be a sequence of $(2k+1)\frac{q}{\gcd(q,(2k+1)j)} + k$ initial terms of c_j^r. Note that s_j^r covers the same pairs as c_j^r. For fixed j, the sum of legths of sequences c_j^r (where $r = 1, 2, \ldots, \gcd(q, (2k+1)j) - 1$) equals $(2k+1)q + k\gcd(q, (2k+1)j)$. Therefore, the sum of lengths of all sequences s_j^r equals

$$(2k+1)q^2 + k\sum_{j=0}^{q-1}\gcd(q, (2k+1)j)$$

$$= \frac{n^2}{2k+1} + k\sum_{j=0}^{q-1}\gcd(q, (2k+1)j).$$

Finally, take s to be concatenation of $(2k+1)$ shortest k-radius sequences over an alphabet of size q (where i-th sequence is over $\{(i, b) : b \in Z_q\}$) with all sequences s_j^r. Clearly by our argument, s covers every pair of elements from A. The length of s is equal to $(2k+1)f_k(q)$ plus the sum of lengths of all s_j^r, which concludes the proof □

The above construction can be used also when n is greater than $q(2k+1)$. Suppose that we extend our alphabet A (of size $(2k+1)q$) by a set B of $n-(2k+1)q$ additional symbols. Note, that there are (at most) $|A|\,|B|$ pairs that contain at least one of the added symbols. Therefore, we can construct a k-radius sequence over $A \cup B$ by appending $2\,|A|\,|B|$ terms to a k-radius sequence over A. Thus, we obtain a corollary:

Corollary 13 ([4, Lemma 2.7])**.** *For all k and n, where $q \leq \frac{n}{2k+1}$ and all divisors of q are greater than k, we have*

$$f_k(n) \leq (2k+1)f_k(\left\lfloor \frac{n}{2k+1} \right\rfloor) + \frac{n^2}{2k+1} + k\sum_{j=0}^{q-1}\gcd(q, (2k+1)j) + 2n(n-(2k+1)q).$$

□

6. Constant k

We will use Corollary 13 to obtain the value of $f_k(n)$ with error $O(n^{1+\epsilon})$. The idea is to choose apropriate value of q for any n, such that two last summands are small enough. Solving the recurrence $g(n) = (2k+1)g(\frac{n}{2k+1}) + \frac{n^2}{2k+1}$ gives us $\frac{n^2}{2k}$, as desired, so we only need to take proper care of additional small terms.

Lemma 14. *Take any k and $\epsilon > 0$. There exists a constatnt c such that for any n we can find $q \leq \frac{n}{2k+1}$ such that all divisors of q are greater than k and*

$$k\sum_{j=0}^{q-1}\gcd(q, (2k+1)j) + 2n(n-(2k+1)q) \leq cn^{1+\epsilon}.$$

Proof. First, we proove the fact that any integer q has no more than $c'q^\epsilon$ divisors, for some constant c'. Indeed, let $q = p_1^{\alpha_1}p_2^{\alpha_2}\dots p_r^{\alpha_r}$ (where p_i are distinct primes arranged in an increasing order). Note that q has exactly $d(q) = (\alpha_1+1)(\alpha_2+1)\dots(\alpha_r+1)$ divisors. We will compare q^ϵ with $d(q)$. Take $q' = p_1^{\alpha_1}p_2^{\alpha_2}\dots p_{r_0}^{\alpha_{r_0}}$, where we choose r_0 such that r_0-th smallest prime p satisfies $p^\epsilon \leq 2$. Define $s = \alpha_1+\alpha_2+\ldots+\alpha_{r_0}$.

Now, there is a constant c' such that $d(q') \leq c'q'^\epsilon$. Indeed, $q' \geq 2^s$ (as every p_i is at least 2) and $d(q') \leq (\frac{s+r_0}{r_0})^{r_0}$ (as the product of r_0 numbers that sum up to $s+r_0$ is largest when those numbers are equal). As r_0 is constant and both expressions are nonzero, we can find c' such that $c'2^s \geq (\frac{s+r_0}{r_0})^{r_0}$. Therefore, we obtain $d(q') \leq c'q^\epsilon$.

Now, consider $q'' = p_{r_0+1}^{\alpha_{r_0+1}}p_{r_0+2}^{\alpha_{r_0+2}}\dots p_r^{\alpha_r}$. By $p_i^\epsilon \leq 2$ we have $p_i^{\epsilon\alpha_i} \geq (\alpha_i+1)$ for $i > r_0$. Therefore, we obtain $d(q'') \leq q''^\epsilon$. As $q = q'q''$ and $d(q) = d(q')d(q'')$, we get $d(q) \leq q^\epsilon$ as desired.

Let d be any divisor of q. We may have $\gcd(q,(2k+1)j) = d$ only if d is a divisor of $(2k+1)j$, which can happen at most $\frac{(2k+1)q}{d}$ times (for $j < q$). Therefore, in the sum $\sum_{j=0}^{q-1}\gcd(q,(2k+1)j)$ every divisor d of q (including 1) contributes at most $d\frac{(2k+1)q}{d} = (2k+1)q$. It follows that

$$\sum_{j=0}^{q-1}\gcd(q,(2k+1)j) \leq (2k+1)c'q^{1+\epsilon}.$$

Take q to be the largest number such that $q \leq \frac{n}{2k+1}$ and q is equal to 1 modulo $k!$. Clearly, the smallest (nontrivial) divisor of q is greater than k and $n-(2k+1)q \leq (2k+1)k!$. We obtain $2n(n-(2k+1)q) \leq 2(2k+1)k!n$. Now, if we take $c = (2k+1)c' + 2(2k+1)k!$ the proof is finished. □

Now, we are ready to proove the main result of this section.

Theorem 15 ([4, Theorem 3.6])**.** *For any fixed k and for every $\epsilon > 0$ we have*

$$f_k(n) = \frac{1}{2k}n^2 + O(n^{1+\epsilon}).$$

Proof. By Corollary 13 and Lemma 14 we get the recurrence $f_k(n) \leq (2k+1)f_k(\left\lfloor\frac{n}{2k+1}\right\rfloor) + \frac{n^2}{2k+1} + cn^{1+\epsilon'}$ (where $\epsilon' = \frac{\epsilon}{2}$). Now, consider the function g defined for all nonnegative reals as

$$g(x) = \begin{cases} (2k+1)g(\frac{x}{2k+1}) + r(x) & \text{for } x \geq 1, \\ 0 & \text{for } 0 \leq x < 1, \end{cases}$$

where $r(x) = \frac{x^2}{2k+1} + cx^{1+\epsilon'}$. By a simple induction on n, we get that $f_k(n) \leq g(n)$ for $n \in \mathbb{N}$.

By expanding the definition, we have

$$g(x) = r(x) + (2k+1)r(\frac{x}{2k+1}) + (2k+1)^2 r(\frac{x}{(2k+1)^2}) + \ldots + (2k+1)^t r(\frac{x}{(2k+1)^t})$$

for $t = \lfloor\log_{2k+1}(x)+1\rfloor$. Therefore

$$g(x) = \sum_{i=0}^{t}(2k+1)^i r(\frac{x}{(2k+1)^i})$$

$$= \sum_{i=0}^{t}(2k+1)^i \frac{1}{2k+1}\frac{x^2}{(2k+1)^{2i}} + \sum_{i=0}^{t}(2k+1)^i c\frac{x^{1+\epsilon'}}{(2k+1)^{i(1+\epsilon')}}$$

$$\leq \frac{x^2}{2k+1}\sum_{i=0}^{\infty}(\frac{1}{2k+1})^i + \sum_{i=0}^{t} c\frac{x^{1+\epsilon'}}{(2k+1)^{i\epsilon'}}$$

$$\leq \frac{x^2}{2k+1}\frac{1}{1-\frac{1}{2k+1}} + tcx^{1+\epsilon'}$$

$$= \frac{x^2}{2k} + c\log_{2k+1}(x)x^{1+\epsilon'}$$

As $\log_{2k+1}(x) = O(x^{\epsilon'})$, we get $f_k(n) \leq g(n) \leq \frac{n^2}{2k} + O(n^{1+\epsilon})$ as desired. □

7. k depending on n

In this section we consider the case when $k = O(n^\alpha)$ for some $\alpha < 1$. In our calculations we will take q to be prime, and easily get a bound for the last two terms, and use it instead of Lemma 14. We rely on the fact that for every integer n there is a prime number within distance $O(n^\delta)$ from n, where $\delta = 0.525$.

Lemma 16 ([4, Lemma 4.1])**.** *There exists x_0 such that for every $n \geq x_0$ the interval $[n - n^\delta, n]$ contains a prime number.*

We begin by reformulating Corollary 13 to get rid of q

Lemma 17. *For all x, n such that $0 < x < 1$ and $n^x \in \mathbb{N}$ we have*

$$f_{n^x}(n) = (2n^x+1)f_{n^x}(\left\lfloor \frac{n}{2n^x+1} \right\rfloor) + \frac{n^2}{2n^x+1} + O(n^{1+\delta(1-x)}).$$

Proof. Take q to be the largest prime that does not exceed $\frac{n}{2n^x+1}$. By Lemma 16 we have $n - (2n^x+1)q \leq O((\frac{n}{2n^x+1})^\delta) = O(n^{\delta(1-x)})$. Moreover, by q being prime that is $O(n^{1-x})$ we get that

$$\sum_{j=0}^{q-1} \gcd(q, (2n^x+1)j) = 2q - 1 = O(n^{1-x}).$$

Therefore, by Corollary 13 we have

$$f_{n^x}(n) \leq (2n^x+1)f_k(\left\lfloor \frac{n}{2n^x+1} \right\rfloor) + \frac{n^2}{2n^x+1} + n^x\sum_{j=0}^{q-1}\gcd(q,(2k+1)j) + 2n(n-(2k+1)q)$$

$$= (2n^x+1)f_{n^x}(\left\lfloor \frac{n}{2n^x+1} \right\rfloor) + \frac{n^2}{2n^x+1} + O(n^{x+(1-x)}) + O(n^{1+\delta(1-x)}).$$

This completes the proof, as $x + (1-x) = 1 \leq 1 + \delta(1-x)$. □

Note that by expanding the recursion in Lemma 17 we get a constant number of summands - depending on $\lfloor \frac{1}{x} \rfloor$, and x is constant. By this observation, we may inductivly apply this result, where the induction in on $\lfloor \frac{1}{x} \rfloor$

Theorem 18. *For all x, n such that $0 < x < 1$ and $n^x \in \mathbb{N}$ we have*

$$f_{n^x}(n) = \frac{n^2}{2n^x} + O(n^{1+\delta(1-x)}).$$

Proof. We will use induction on $\lfloor \frac{1}{x} \rfloor$ to prove that

$$f_{n^x}(n) \leq \sum_{i=1}^{\frac{1}{x}} \frac{n^2}{(2n^x+1)^i} + O(n^{1+\delta(1-x)})$$

. For the base of induction, note that $\frac{n}{2n^x+1} \leq n^{1-x}$ and for $\lfloor \frac{1}{x} \rfloor = 1$ we have $n^x \leq n^{1-x}$. As $f_a(b) = b$ when $a \geq b$, by Lemma 17 we obtain $f_{n^x}(n) = (2n^x + 1)\frac{n}{2n^x+1} + \frac{n^2}{2n^x+1} + O(n^{1+\delta(1-x)}) = \frac{n^2}{2n^x+1} + O(n^{1+\delta(1-x)})$ as desired.

To perform the induction step, we shall rewrite the statement of Lemma 17 as

$$f_{n^x}(n) \leq (2n^x + 1) f_{n'^{x'}}(n') + \frac{n^2}{2n^x+1} + O(n^{1+\delta(1-x)})$$

for $n' = \left\lfloor \frac{n}{2n^x+1} \right\rfloor$ and $1 > x' \geq \frac{x}{1-x}$ (chosen such that $n'^{x'} = n^x$). Clearly $\lfloor \frac{1}{x'} \rfloor \leq \lfloor \frac{1}{x} - 1 \rfloor$, so by the induction assumption we obtain

$$f_{n^x}(n) \leq (2n^x + 1)(\sum_{i=1}^{\frac{1}{x'}} \frac{n'^2}{(2n'^{x'}+1)^i} + O(n'^{1+\delta(1-x')})) + \frac{n^2}{2n^x+1} + O(n^{1+\delta(1-x)}).$$

By rearranging terms and using definitions of n' and x' we get

$$f_{n^x}(n) \leq \frac{n^2}{2n^x+1} + (2n^x+1)(\frac{n}{2n^x+1})^2 \sum_{i=1}^{\frac{1}{x'}} \frac{1}{(2n^x+1)^i}$$

$$+(2n^x+1)O((\frac{n}{2n^x+1})^{1+\delta(1-x')}) + O(n^{1+\delta(1-x)})$$

$$= \sum_{i=1}^{\frac{1}{x'}+1} \frac{n^2}{(2n^x+1)^i} + O((2n^x+1)^{1-1-\delta(1-x')} n^{1+\delta(1-x)} + n^{1+\delta(1-x)}).$$

We have $\frac{1}{x'} + 1 \leq \frac{1}{x}$ and $\delta(1-x') > 0$, therefore the proof of the claim is finished.

As the infinite sum $\sum_{i=1}^{\infty} \frac{n^2}{(2n^x+1)^i}$ is equal $\frac{n^2}{2n^x+1} \frac{1}{1-\frac{1}{2n^x+1}} = \frac{n^2}{2n^x}$, our claim proves the desired result. □

The assumption that $n^x \in \mathbb{N}$ in Theorem 18 is only technical. For general case, we can choose $x' < x$ such that $n^{x'} = \lfloor n^x \rfloor$ and apply Theorem 18. As $n^{x-x'}$ is close to 1, we obtain the following corollary:

Corollary 19. *For all x such that $0 < x < 1$ we have*

$$f_{\lfloor n^x \rfloor}(n) = \frac{n^2}{2\lfloor n^x \rfloor} + O(n^{1+\delta(1-x)}).$$

□

8. Conclusions

The study of k-radius sequences was motivated by the problem of comparing all pairs from the set of large objects, where local memory is managed by FIFO (first in, first out) rule.

It turned out that the restriction to FIFO is not too expensive. Indeed, the lower bound from Lemma 5 can be easily proved for any other model of memory management and we showed that this bound is already (asymptotically) attained by k-radius sequences in case when k is significantly smaller than the alphabet.

In Theorem 15 we stated that for a constatnt k we have $f_k(n) = \frac{n^2}{2k} + O(n^{1+\epsilon})$, where the ϵ emerged from estimating the number of divisors of a natural number. We believe that the last term can be improved to $O(n)$ by a better choice of the value q (for Lemma 14) or a more involved construction. In fact, in [2] it is proved for some specific values of n and k.

The bound in Theorem 18 ($\frac{n^2}{2k} + O(O(n^{1+\delta(1-x)}))$) depends on a constant δ, that comes from Lemma 16. However, for n such that $\frac{n}{2k+1}$ is close to a prime, we can do better. As the problem does not seem to directly depend on whether n is prime, it gives some hope for improvement in the general case.

Note that in case when k is a linear function of n, our constructions would give the leading term in bound on $f_k(n)$ that is larger than $\frac{n^2}{2k}$ (as it is $O(n)$). Therefore, it is an open problem to determine the (asymptotically) exact value of $f_{cn}(n)$, or at least provide a lower bound that is better than $\frac{n^2}{2k}$.

Blackburn and McKee [2] provide other (not explicit) construction of k-radius sequences, that relies on logarithms of length k, where a logarithm of length k is a map $f : [k] \to \mathbb{Z}_k$ that satisfies $f(ij) = f(i) + f(j)$ whenever $ij \leq k$. In other words, logarithm of length k corresponds to completing a partialy filled multiplication table of numbers $1, 2, \ldots, k$ so as to obtain a table of a cyclic group (of order k). It is known that such objects does not exist for some values of k, the smallest counterexample being $k = 195$.

A logarithm of length k corresponds to a coloring of a graph $\mathbb{B}_k$ with elements of $\mathbb{Z}_k$, where $\mathbb{B}_k$ is defined as follows. The vertices of $\mathbb{B}_k$ are all natural numbers and there is an edge $\{ai, aj\}$ for all $a \in \mathbb{N}$ and $i, j \in [k]$. A conjecture of B. Bosek states that the graph $\mathbb{B}_k$ can be colored with k colors for all k, and such coloring may be thought of as an object more general than logarithm of length k. The conjecture is still open for $k = 195$.

It is intriguing how properties of the graph $\mathbb{B}_k$ relate to k-radius sequences and it seems an interesting research direction.

References

[1] S.R. Blackburn, The existence of k-radius sequences, Journal of Combinatorial Theory, Ser. A, 119 (2012), 212–217.

[2] S.R. Blackburn, J.F. McKee, Constructing k-radius sequences, Mathematics of Computation, to appear.

[3] J. Jaromczyk, Z. Lonc, Sequences of radius k: how to fetch many huge objects into small memory for pairwise computations, ISAAC 2004, HongKong, Lecture Notes in Computer Science 3341 (2004), 594-Ũ605.

[4] J. Jaromczyk, Z. Lonc, M. Truszczynski, Constructions of asymptotically shortest k-radius sequences, Journal of Combinatorial Theory, Ser. A, 119 (2012), 731-Ũ746.