ON THE MINIMAL DISTANCE OF A POLYNOMIAL CODE

PÉTER PÁL PACH AND CSABA SZABÓ

1. Introduction

For two finite subsets of the positive integers, A and B let $A * B = \{ab \mid a \in A,\, b \in B \text{ and } ab \text{ occurs odd many times in } A \cdot B\}$. In other words, if $A = \{a_1, \dots, a_k\}$, then $A * B = a_1 B \Delta \cdots \Delta a_k B$, where Δ denotes the symmetric difference. For a positive integer m let $\underline{m} = \{1, 2, \dots, m\}$.

Conjecture 1. If n, k are positive integers, then $|\underline{n} * \underline{k}| \geq n$.

For an arbitrary finite subset $A \subset \mathbb{N}$ it was proved that $|\underline{m} * A| \geq \pi(m) + 1$, where $\pi(x)$ is the prime counting function, and the following conjecture was formulated ([5]):

Conjecture 2. Let n be a positive integer and $K \subset \mathbb{N}$ be a finite set of integers. Then $|\underline{n} * K| \geq n$.

These purely number theoretical problems originate in the theory of near-ring codes. A near-ring can be described as a ring, where the addition is not necessarily commutative and only one of the distributive laws is required. A typical example is the near-ring of polynomials, where the addition is the usual polynomial addition, and multiplication is the composition of the polynomials. In this example the addition is commutative and only the right distributive law holds. Near-rings play an important role in combinatorics: They are used to construct block designs that give rise to efficient error correcting codes. For more information on these codes see [2], [3] and [4]. A special and very interesting near-ring code is defined in the following way: Let $f \in \mathbb{Z}_2[x]$ be a polynomial and $C(f, k)$ the code generated (as a subspace) by the polynomials $f = f \circ x, f \circ x^2, \dots, f \circ x^k$. For $f = x + x^2 + \cdots + x^n$ a typical codeword is

$$\sum_{i \in K} f \circ x^i = \sum_{j \in K * \underline{n}} x^j,$$

where K is a finite subset of $\underline{k}$. As $C(f, k)$ is a linear code, its minimal distance is equal to the minimal weight of any nonzero codeword. Hence the minimum distance of $C(f,k)$ is the minimal value of $|\underline{n}*K|$ for some $K \subseteq \underline{k}$.

In this paper we settle Conjecture 1, and prove that for arbitrary $n \in \mathbb{N}$ and finite set $K \subset \mathbb{N}$ we have $|\underline{n} * K| \geq c \cdot \dfrac{n}{\log^{0.223} n}$ for some $c > 0$. Note that the minimal distance in $C(f,k)$ depends heavily on f. If, for example, we start with $f(x) = x + x^2 + x^4 + \cdots + x^{2^k}$, then $f \circ x + f \circ x^2 = x + x^{2^{k+1}}$, hence the minimal distance of the corresponding code is 2.

The natural logarithm will be denoted by log through the whole paper.

2. The general case

Let us denote by $g(n)$ the minimal size of the set $\underline{n} * K$, where K is a finite subset of the positive integers. In [5] it is proved that $g(n) \geq \pi(n) + 1$. In this section we improve this lower bound and prove that $g(n) \geq c \cdot \dfrac{n}{\log^{0.223} n}$ for some $c > 0$. The proof is based on the following lemma:

Proposition 1. *For every positive integer n*

$$g(n) \geq \sum_{p \leq n} g\left(\lfloor n/p^{\alpha_p} \rfloor\right),$$

where the sum goes over the primes less than n, and α_p is the largest integer such that $p^{\alpha_p} \leq n$.

Proof. Let $p \leq n$ be a prime and $K_p \subseteq K$ the subset of K containing the elements that are divisible by the largest power of p occuring as divisor of some element of K (possibly $p^0 = 1$). Similarly, let $\underline{n}_p \subseteq \underline{n}$ be the set of elements of $\underline{n}$ that are divisible by p^{α_p}. Note that $\underline{n}_p$ is never empty. By the maximality of the exponents of p in K_p and $\underline{n}_p$, for any $a \in \underline{n}_p$, $b \in K_p$ and $c \in \underline{n}$, $d \in K$ if $ab = cd$, then $c \in \underline{n}_p$ and $d \in K_p$ hold. We prove that for $p < q \leq n$ different primes $\underline{n}_p \cdot K_p$ and $\underline{n}_q \cdot K_q$ are disjoint. If for some $a \in \underline{n}$ and $b \in K$ we have $ab \in \underline{n}_p \cdot K_p \cap \underline{n}_q \cdot K_q$, then $a \in \underline{n}_p \cap \underline{n}_q$. Thus $a = pqd'$, and $\bar{a} = p^2 d' < a$ is in $\underline{n}$. The exponent of p in $\bar{a}$ is larger than the one in a, which is contradiction. Hence, $\underline{n} * K$ contains the disjoint union of the sets $\underline{n}_p \cdot K_p$ for $p \leq n$, so

$$|\underline{n} * K| \geq \sum_{p \leq n} |\underline{n}_p * K_p|. \tag{1}$$

As $p^{\alpha_p} \le n < p^{\alpha_p+1}$, clearly, $\underline{n}_p = \{p^{\alpha_p}, 2p^{\alpha_p}, \dots, \lfloor n/p^{\alpha_p}\rfloor p^{\alpha_p}\}$, where $\lfloor n/p^{\alpha_p}\rfloor < p$. Dividing by p^{α_p}, we obtain that $|\underline{n}_p * K_p| = |\underline{\lfloor n/p^{\alpha_p}\rfloor} * K_p|$, thus by the definition of g we get

$$|\underline{n}_p * K_p| = |\underline{\lfloor n/p^{\alpha_p}\rfloor} * K_p| \ge g(\lfloor n/p^{\alpha_p}\rfloor).$$

By (1) we have

$$g(n) \ge \sum_{p\le n} g\left(\lfloor n/p^{\alpha_p}\rfloor\right),$$

and this is what we wanted to prove. □

Theorem 2. *For every $\lambda > \lambda_0$ there exists a $c = c(\lambda) > 0$ such that for every $n > 1$*

$$g(n) \ge c \cdot \frac{n}{\log^\lambda n},$$

where λ_0 satisfies $\int\limits_0^1 \left(\frac{2}{y}\right)^{\lambda_0} \frac{1}{2-y} dy = 1$. *Note that $\lambda_0 \sim 0.2223...$*

Proof. Fix $1 > \lambda > \lambda_0$. We claim that there exists some $c > 0$ such that the inequality

$$g(n) \ge c \cdot \frac{n}{\log^\lambda n} \tag{2}$$

holds for every $n > 1$. The proof is by induction on n. First we discuss the induction step. Assume that (2) holds for $n < m$. Now, we show that it holds for $n = m$, as well. The value of c will be chosen later. By Proposition 1 and the induction hypothesis:

$$\begin{aligned}
g(m) &\ge \sum_{\sqrt{m}<p\le m} g\left(\lfloor m/p\rfloor\right) \ge \sum_{\sqrt{m}<p<m/2} c\cdot \frac{\lfloor m/p\rfloor}{\log^\lambda(\lfloor m/p\rfloor)} \ge \\
&\ge \sum_{\sqrt{m}<p<m/2} c\cdot \frac{\lfloor m/p\rfloor}{\log^\lambda(\lfloor m/p\rfloor)} \ge \sum_{\sqrt{m}<p<m/2} c\cdot \frac{m/p-1}{\log^\lambda(\lfloor m/p\rfloor)} = \\
&= \sum_{\sqrt{m}<p<m/2} c\cdot \frac{m/p}{\log^\lambda(\lfloor m/p\rfloor)} - \sum_{\sqrt{m}<p<m/2} c\cdot \frac{1}{\log^\lambda(\lfloor m/p\rfloor)}.
\end{aligned} \tag{3}$$

In [7] it is proved that $\pi(m) < \frac{1.25506m}{\log m}$ for every $m > 1$, hence $\pi(m/2) - \pi(\sqrt{m}) \le \pi(m) < 1.5 \cdot \dfrac{m}{\log m}$. For the second term of the last line of (3) we obtain:

(4)

$$\sum_{\sqrt{m}<p<m/2} c\cdot \frac{1}{\log^\lambda(\lfloor m/p\rfloor)} \le \sum_{\sqrt{m}<p<m/2} c\cdot \frac{1}{(\log 2)^\lambda} \le 1.5\cdot \frac{m}{\log m}\cdot \frac{c}{\log 2} = o\left(\frac{m}{\log^\lambda m}\right),$$

since $\lambda < 1$.

Now we estimate the main term. By Mertens' theorem, there exists a constant M such that $\sum_{p\le x} \frac{1}{p} = \log\log x + M + o(1)$. Hence, for every $\varepsilon > 0$ there exists $B = B(\varepsilon)$ such that for $B \le a \le b$

$$\left| \sum_{a<p<b} \frac{1}{p} - \log\log b + \log\log a \right| < \varepsilon \tag{5}$$

holds. For $m > 2^{2K}$ we have $m^{\frac{1}{2}+\frac{K-1}{2K}} < m/2$. Applying (5) to the interval $I_h = (m^{\frac{1}{2}+\frac{h-1}{2K}}, m^{\frac{1}{2}+\frac{h}{2K}}]$, where h is an integer satisfying $1 \le h \le K-1$ we obtain that

$$\sum_{p\in I_h} \frac{1}{p} > \log \frac{K+h}{K+h-1} - \varepsilon. \tag{6}$$

If $p \in I_h$, then $\log^\lambda(m/p) \le \log^\lambda(m)(\frac{K-h+1}{2K})^\lambda$. Substituting into the main term of the last line of (3), omitting the integer parts and rearranging we get that

$$\begin{aligned} \sum_{\sqrt{m}<p<m/2} c\cdot\frac{m/p}{\log^\lambda(\lfloor m/p\rfloor)} &\ge cm \sum_{\sqrt{m}<p<m/2} \frac{1/p}{\log^\lambda(m/p)} \ge \\ &\ge \frac{cm}{\log^\lambda m} \sum_{h=1}^{K-1}\sum_{p\in I_h} \left(\frac{2K}{K-h+1}\right)^\lambda \cdot \frac{1}{p} \ge \\ \ge \frac{cm}{\log^\lambda m}\left(\sum_{h=1}^{K-1} \left(\frac{2K}{K-h+1}\right)^\lambda \log\frac{K+h}{K+h-1} \right.&\left. - \varepsilon \sum_{h=1}^{K-1}\left(\frac{2K}{K-h+1}\right)^\lambda \right). \end{aligned} \tag{7}$$

Now we show that there exists some K such that

$$S_K = \sum_{h=1}^{K-1} \left(\frac{2K}{K-h+1}\right)^\lambda \log\frac{K+h}{K+h-1} > 1. \tag{8}$$

Let $f_K(y) = \left(\frac{2}{y}\right)^\lambda K\cdot\log\left(1+\frac{1}{K(2-y)}\right)$ and $f(y) = \left(\frac{2}{y}\right)^\lambda \cdot \frac{1}{2-y}$. The sequence of functions f_K converges to f. Then

$$S_K = \frac{f_K(\frac{1}{K}) + f_K(\frac{2}{K}) + \cdots + f_K(\frac{K}{K})}{K} - \frac{f_K(\frac{1}{K})}{K}.$$

Let

$$T_K = \frac{f(\frac{1}{K}) + f(\frac{2}{K}) + \cdots + f(\frac{K}{K})}{K}.$$

As $1 > \lambda > \lambda_0$, the Riemann-sum T_k converges to $\int_0^1 f > 1$. As $f_K(\frac{1}{K})/K$ converges to 0, it is easy to see that $S_K - T_K$ converges to 0. Hence we can fix a K such that $S_K > 1$. Now, we can choose some $\varepsilon > 0$ such that

$$\eta = \sum_{h=1}^{K-1} \left(\frac{2K}{K-h+1}\right)^\lambda \log\frac{K+h}{K+h-1} - 1 - \varepsilon \sum_{h=1}^{K-1} \left(\frac{2K}{K-h+1}\right)^\lambda > 0.$$

According to (4) there exists some R such that if $R < m$, then

$$\sum_{\sqrt{m}<p<m/2} c \cdot \frac{1}{\log^\lambda(\lfloor m/p \rfloor)} \leq \eta \cdot c \cdot \frac{m}{\log^\lambda m}.$$

By (3) and (7) we obtain that $g(m) \geq c \cdot \dfrac{m}{\log^\lambda m}$ holds. If we choose $c > 0$ such that (2) holds for $n \leq \max(2^{2K}, B^2(\varepsilon), R)$, then (3) is gained.

□

3. The case $K = \underline{k}$

In this section we prove Conjecture 1. We distinguish cases according to how large is k compared to n. The conjecture is true for $k \leq 8$. ([5])

3.1. **Case 1:** $9 \leq k \leq 1.34 \cdot \log n$

We show that in this case the number of elements that occur exactly once in the product $\underline{n} \cdot \underline{k}$ is at least n. We shall need the following two observations.

Lemma 3. *Let $n/2 < a \leq n$ and $b \in \underline{k}$ such that a is relatively prime to every number less than k. Then ab occurs once in $\underline{n} \cdot \underline{k}$.*

Proof. Let us assume that $a_1, a_2 \in \underline{n}$ and $b_1, b_2 \in \underline{k}$ satisfy the conditions of the lemma, and $a_1 b_1 = a_2 b_2$. Now, $a_1 | a_2 b_2$ and a_1 and b_2 are relatively prime, hence $a_1 | a_2$. As $a_1 > n/2$ we have $2a_1 > n \geq a_2$, thus $a_1 = a_2$, which implies $b_1 = b_2$. □

Lemma 4. *If $k \geq 14$, then* $\displaystyle\prod_{p \leq k} \left(1 - \frac{1}{p}\right) \geq \frac{0.5}{\log k}$.

Proof. In [7] it is shown that for $k > 1$

$$\frac{e^{-\gamma}}{\log k} \left(1 - \frac{1}{\log^2 k}\right) \leq \prod_{p \leq k} \left(1 - \frac{1}{p}\right),$$

where γ is the Euler constant. For $k > 21$ by using the monotonicity of the logarithm function and $e^{-\gamma} > 0.56$ we get that

$$\frac{e^{-\gamma}}{\log k}\left(1-\frac{1}{\log^2 k}\right) \geq \frac{0.56}{\log k}\left(1-\frac{1}{\log^2 22}\right) > \frac{0.5}{\log k}.$$

For $14 \leq k \leq 21$ it is enough to check the statement when $k = 14$, 17 and 19. For these numbers the values of $(\log k)\cdot \prod_{p\leq k}\left(1-\frac{1}{p}\right)$ are 0.506, 0.511 and 0.503, respectively, hence the statement holds. □

Proposition 5. *Let* $9 \leq k \leq 1.34 \cdot \log n$. *Then* $|\underline{n} * \underline{k}| \geq n$.

Proof. We show that there are at least n products satisfying the conditions of Lemma 3. For this we need to estimate the number of integers between $n/2$ and n that are not divisible by a prime less than k. This number will be denoted by D. By the inclusion-exclusion principle

$$\text{(9)}\quad D = n - \lfloor n/2\rfloor + \\ +\sum_{h=1}^{r}(-1)^h \sum_{1\leq i_1<\ldots<i_h\leq r}\left(\left\lfloor\frac{n}{p_{i_1}\cdots p_{i_h}}\right\rfloor - \left\lfloor\frac{n/2}{p_{i_1}\cdots p_{i_h}}\right\rfloor\right),$$

where $\pi(k) = r$ and $p_1, \ldots, p_r$ are the primes up to k. Applying $x-1 < \lfloor x\rfloor \leq x$ to all 2^{r+1} terms of the right side we get that

$$\text{(10)}\quad D \geq n - n/2 + \\ +\sum_{h=1}^{r}(-1)^h \sum_{1\leq i_1<\ldots<i_h\leq r}\left(\frac{n}{p_{i_1}\cdots p_{i_h}} - \frac{n/2}{p_{i_1}\cdots p_{i_h}}\right) - 2^r = \\ = \frac{n}{2}\prod_{p\leq k}\left(1-\frac{1}{p}\right) - 2^r.$$

If $k \geq 14$, Lemma 4 applies, and

$$D \geq \frac{n}{2}\prod_{p\leq k}\left(1-\frac{1}{p}\right) - 2^r \geq \frac{0.25n}{\log k} - 2^r.$$

As $k \leq 1.34\log n$, for $k \geq 14$ we have the estimation

$$2^r = 2^{\pi(k)} \leq 2^{k/2} \leq \frac{1}{100\log k}\cdot e^{\frac{k}{1.34}} \leq \frac{n}{100\log k}.$$

Hence, $D \geq \dfrac{0.24n}{\log k}$. Using Lemma 3 we obtain $|\underline{n} * \underline{k}| \geq Dk$. The function $x/\log x$ is monotone increasing on $[1, \infty)$, thus

$$|\underline{n} * \underline{k}| \geq Dk \geq \frac{0.24k}{\log k} n \geq \frac{0.24 \cdot 14}{\log 14} n > n.$$

For $9 \leq k \leq 13$ we have

$$|\underline{n} * \underline{k}| \geq Dk \geq \left(\frac{n}{2} \prod_{p \leq k} \left(1 - \frac{1}{p}\right) - 2^{\pi(k)} \right) k.$$

For $10 \leq k \leq 13$ it is obtained by calculation that the right hand side is greater than n if $n \geq e^{k/1.34}$. For $k = 9$ the inequality holds if $n > 5040$. By brute force the statement can be checked for $k = 9$ and $n \leq 5040$. Thus we obtained $|\underline{n} * \underline{k}| > n$.

□

3.2. **Case 2:** $1.34 \cdot \log n \leq k \leq n - \dfrac{0.22 \cdot n}{\log n}$ and $n \geq 1410$.

Let $k_1 = \max(k, n/7)$ and $k_1 < p \leq n$ a prime. As $k < p$, the set of elements of $\underline{n} * \underline{k}$, which are divisible by p is $\{p, 2p, \ldots, \lfloor n/p \rfloor p\} * \underline{k}$. This set has the same cardinality as the set $\underline{\lfloor n/p \rfloor} * \underline{k}$. Now, $\lfloor n/p \rfloor \leq 6$, hence $|\underline{\lfloor n/p \rfloor} * \underline{k}| \geq k$. It is easy to see that for $p > q > n/7$ an element of $\underline{n} * \underline{k}$ cannot be divisible by both p and q. Hence, $|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k_1))k$.

At first, suppose that $k \leq n/7$. By a theorem of Dusart [1] for $x \geq 17$

$$\frac{x}{\log x} \leq \pi(x) \leq \frac{x}{\log x} \left(1 + \frac{1.2762}{\log x}\right)$$

holds. Hence, $\pi(n) - \pi(n/7) \geq 0.749 \cdot \dfrac{n}{\log n}$ for $n \geq 1410$. As $1.34 \cdot \log n \leq k$, we have

$$|\underline{n} * \underline{k}| \geq 1.34 \cdot 0.749 \cdot n > n.$$

Secondly, let us consider the case when $n/7 < k \leq n/2$. As $\pi(n) - \pi(n/2) \geq 7$,

$$|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k_1))k > 7 \cdot n/7 = n.$$

Finally, let $n/2 < k < n - \dfrac{0.22 \cdot n}{\log n}$. Then by the estimates in [1] and [6] there are at least two primes between k and n if $n > 90000$. It can be checked that this also holds for $n > 1410$. Thus

$$|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k))k \geq 2(n/2) = n.$$

We continue with the case when k is "large", that is, $n-\frac{0.4\cdot n}{\log n+1.02}\le k$. By calculation we have $n-\frac{0.4\cdot n}{\log n+1.02}\le n-\frac{0.22\cdot n}{\log n}$ for $n\ge 4$.

3.3. Case 3. $n-\dfrac{0.4\cdot n}{\log n+1.02}\le k\le n$ and $n>5000$.

If $k=n$, then $\underline{k}\cdot\underline{n}=\{1,\dots,n\}\cdot\{1,\dots,n\}$. If $a\neq b$, then pairing ab with ba only the products of the form $a\cdot a$ are left, hence $\underline{n}*\underline{k}=\{1^2,2^2,\dots,n^2\}$. Thus

$$|\underline{n}*\underline{k}|=n.$$

Assume now that $k<n$. Then

$$\begin{aligned}|\underline{n}*\underline{k}|&=|(\underline{k}*\underline{k})\Delta((\underline{n}\setminus\underline{k})*\underline{k})|=\\ &=|\underline{k}*\underline{k}|+|(\underline{n}\setminus\underline{k})*\underline{k}|-2|(\underline{k}*\underline{k})\cap((\underline{n}\setminus\underline{k})*\underline{k})|.\end{aligned}\tag{11}$$

For the first term on the right side of (11) we have

$$|\underline{k}*\underline{k}|=|\{1^2,2^2,\dots,k^2\}|=k.\tag{12}$$

Lemma 6. *For the second term of (11) we have*

$$|(\underline{n}\setminus\underline{k})*\underline{k}|\ge 2k-n.\tag{13}$$

Proof. We use the following observation: If

$$i\le\frac{k}{n-k}\quad\text{ and }\quad k+1\le j\le n,$$

then ij appears exactly once in $(\underline{n}\setminus\underline{k})\cdot\underline{k}$, so $ij\in(\underline{n}\setminus\underline{k})*\underline{k}$. Let us assume that $ij=i'j'$ such that $1\le i'\le k$ and $k+1\le j'\le n$. If $i=i'$, then $j=j'$. If $i'<i$, then $1\le i'\le\dfrac{k}{n-k}$ and $k+1\le j'\le n$. Now, changing the roles of (i,j) and (i',j') we may assume that $i<i'$. As $ij=i'j'$, we have $\dfrac{i}{i'}=\dfrac{j'}{j}$ and

$$\frac{i}{i'}\le\frac{i}{i+1}\le\frac{\frac{k}{n-k}}{\frac{k}{n-k}+1}=\frac{k}{n}<\frac{k+1}{n}\le\frac{j'}{j},$$

which is a contradiction. For $(\underline{n}\setminus\underline{k})*\underline{k}$ we obtain that

$$\begin{aligned}|(\underline{n}\setminus\underline{k})*\underline{k}|&\ge\left\lfloor\frac{k}{n-k}\right\rfloor(n-k)\ge\\ &\ge\left(\frac{k}{n-k}-1\right)(n-k)=k-(n-k)=2k-n.\end{aligned}\tag{14}$$

□

Now, we focus on the third term of (11).

Lemma 7. *For the third second term of (11)*

$$|(\underline{k} * \underline{k}) \cap ((\underline{n} \setminus \underline{k}) * \underline{k})| \leq 0.431 \cdot k. \tag{15}$$

holds.

Proof. It is enough to show that among the numbers 1^2, 2^2, ..., k^2 at most $0.431k$ many has a divisor in the interval $[k+1, n]$. Let $k+1 \leq m \leq n$ and $m = a_m b_m^2$, where b_m^2 is the largest square divisor of m. Since a_m is squarefree, $m | i^2$ if and only if $a_m b_m | i$. Let S denote the following upper bound of the number of elements of the set $\{1^2, 2^2, \ldots, k^2\}$ which have a divisor in $[k+1, n]$:

$$S = \sum_{m=k+1}^{n} \left\lfloor \frac{k}{a_m b_m} \right\rfloor \leq \sum_{m=k+1}^{n} \frac{k}{a_m b_m} = k \sum_{m=k+1}^{n} \frac{b_m}{m}.$$

Recall that $m = a_m b_m^2$, where a_m is squarefree. Now, summing by $j = b_m \leq \sqrt{m}$:

$$S = k \sum_{j=1}^{\lfloor \sqrt{n} \rfloor} \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n, \\ |\mu(m/j^2)|=1}} \frac{j}{m} \leq k \sum_{j=1}^{\lfloor \sqrt{n} \rfloor} j \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{1}{m}.$$

Rewrite $S = k(S_1 + S_2)$, where

$$S_1 := \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{1}{m} \quad \text{and} \quad S_2 := \sum_{j=\lfloor \sqrt{n}/2 \rfloor + 1}^{\lfloor \sqrt{n} \rfloor} j \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{1}{m}.$$

First, we give an upper bound for S_1.

Lemma 8.

$$S_1 \leq \left(\frac{\log n}{2} + 0.31 \right) (\log n - \log k) + \frac{n + 2\sqrt{n}}{8k}. \tag{16}$$

Proof. Let $r_j = \left\lceil \frac{k+1}{j^2} \right\rceil$ and $s_j = \left\lfloor \frac{n}{j^2} \right\rfloor$. Then

$$S_1 = \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j \sum_{l=r_j}^{s_j} \frac{1}{l j^2} = \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \sum_{l=r_j}^{s_j} \frac{1}{l}. \tag{17}$$

The function $\frac{1}{x}$ is a nonnegative decreasing function on $(0,\infty)$, hence we can estimate the inside sum by

$$\sum_{l=r_j}^{s_j} \frac{1}{l} \leq \int_{r_j}^{s_j} 1/x + \frac{1}{r_j} = \log s_j - \log r_j + \frac{1}{r_j}.$$

As $\dfrac{k}{j^2} \leq r_j$ and $s_j \leq \dfrac{n}{j^2}$ we have

$$\log s_j - \log r_j = \log \frac{s_j}{r_j} \leq \log \frac{n/j^2}{k/j^2} = \log n - \log k.$$

Substituting into (17) we obtain

$$(18) \quad S_1 \leq \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j}\left(\log s_j - \log r_j + \frac{1}{r_j}\right) \leq \\ \leq \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j}\left(\log n - \log k + \frac{j^2}{k}\right).$$

Since

$$(19) \quad \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \leq \log\lfloor \sqrt{n}/2 \rfloor + 1 \leq \frac{\log n}{2} - \log 2 + 1 \leq \frac{\log n}{2} + 0.31$$

and

$$(20) \quad \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j = \frac{\lfloor \sqrt{n}/2 \rfloor \cdot (\lfloor \sqrt{n}/2 \rfloor + 1)}{2} \leq \frac{n+2\sqrt{n}}{8},$$

from the inequalities (18), (19), (20) we get (16). □

Now we give an upper bound for S_2.

Lemma 9.
(21)

$$S_2 \leq \left(1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \cdot \frac{n-k}{2\sqrt{k}} \cdot \frac{\sqrt{n}}{k} + \frac{3\sqrt{n}}{k} < 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k}.$$

Proof. Recall that

$$(22) \quad S_2 = \sum_{j=\lfloor \sqrt{n}/2 \rfloor + 1}^{\lfloor \sqrt{n} \rfloor} \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{j}{m}.$$

In (22) for every j we have

$$n \geq j^2 \geq (\lfloor \sqrt{n}/2 \rfloor + 1)^2 > \frac{n}{4}.$$

Hence $m = j^2$ or $2j^2$ or $3j^2$. As $k < m \leq n$, for $m = ij^2$ $(i = 1, 2, 3)$ we get

$$\sqrt{\frac{k}{i}} < j \leq \sqrt{\frac{n}{i}} \quad \text{and} \quad \frac{j}{m} \leq \frac{\sqrt{n}}{k}.$$

For fixed i, the number of j such that $m = ij^2$ is at most:

$$\left\lceil \frac{\sqrt{n} - \sqrt{k}}{\sqrt{i}} \right\rceil = \left\lceil \frac{1}{\sqrt{i}} \cdot \frac{n-k}{\sqrt{n} + \sqrt{k}} \right\rceil \leq \frac{1}{\sqrt{i}} \cdot \frac{n-k}{2\sqrt{k}} + 1,$$

thus

$$S_2 \leq \left(1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \cdot \frac{n-k}{2\sqrt{k}} \cdot \frac{\sqrt{n}}{k} + \frac{3\sqrt{n}}{k} < 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k},$$

and this is what we wanted to show. □

Summarizing the results, from (16) and (21) we obtain:

$$(23) \quad S = k(S_1 + S_2) \leq$$
$$\leq k\left\{\left(\frac{\log n}{2} + 0.31\right)(\log n - \log k) + \frac{n + 2\sqrt{n}}{8k} + 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k}\right\}.$$

We assumed that $n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k$ and $n \geq 5000$. By using the inequality $e^{-x} < \frac{1}{1+x}$ we obtain that $ne^{-\frac{0.2}{\frac{\log n}{2} + 0.31}} < n \cdot \frac{1}{1 + \frac{0.2}{\frac{\log n}{2} + 0.31}} = n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k$. As $n \geq 5000$, we have that $\frac{k}{n} > 0.958$. By easy calculation from these inequalities the following ones can be deduced:

$$(24) \qquad \left(\frac{\log n}{2} + 0.31\right)(\log n - \log k) < 0.2,$$

$$(25) \qquad \frac{n + 2\sqrt{n}}{8k} < 0.135,$$

$$(26) \qquad 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k} < 0.096.$$

Adding (24), (25) and (26) using (23) we arrive at:

$$S \leq k\,(0.2 + 0.135 + 0.096) = 0.431 \cdot k. \tag{27}$$

Then from inequalities (12), (13) and (15) in case $k/n > 0.958$ we get

$$|\underline{k} * \underline{n}| \geq k + 2k - n - 2S \geq 2.138 \cdot k - n > n,$$

thus we proved the statement in Case 3 as well. □

We proved the statement for all pairs n, k where $n \geq 5000$. Cases $k \leq n \leq 5000$ can be checked by brute force.

References

[1] P. Dusart: *The k^{th} prime is greater than $k(lnk + lnlnk - 1)$ for $k > 2$*, Math. Comp., 68:225 (January 1999), 411-415.

[2] R. Eggetsberger: *On Constructing Codes from Planar Nearrings*, http://www.algebra.uni-linz.ac.at/Nearrings/nrcodes.html

[3] G. Pilz: *Near-Rings*, North-Holland Publishing Company (1983), ISBN: 0 7204 0566 1

[4] G. Pilz: *Near-rings: What they are and what they are good for*, http://www.algebra.uni-linz.ac.at/Nearrings/what-are.html

[5] G. Pilz: *On polynomial near-ring codes*, Contributions to general algebra 8, Verlag Hölder-Pichler-Tempsky, Wien (1992), 233-238.

[6] G. Robin: *Estimation de la fonction de tschebyshef theta sur le k-ieme nombre premier et grandes valeurs de la fonction $w(n)$, nombre de diviseurs premiers de n*, Acta. Arith., 42:4 (1983), 367-389.

[7] J. B. Rosser, L. Schoenfeld: *Approximate formulas for some functions of prime numbers*, Ill. Journ. Math. 6 (1962), 64-94.

E-mail address: ppp24@cs.elte.hu
E-mail address: csaba@cs.elte.hu

Eötvös Loránd University, Department of Algebra and Number Theory, 1117 Budapest, Pázmány Péter sétány 1/c, Hungary