ELSEVIER

Discrete Mathematics 214 (2000) 123–133

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

On the number of ordered factorizations of natural numbers

Benny Chor[a,*,1], Paul Lemke[b], Ziv Mador[c,2]

[a] *Department of Computer Science, Technion, Haifa 32000, Israel*
[b] *Center for Communications Research, Thanet Road, Princeton, NJ 08540, USA*
[c] *Department of Computer Science, Technion, Haifa 32000, Israel*

Received 15 May 1997; revised 26 February 1998; accepted 16 November 1998

Abstract

We study the number of ways to factor a natural number n into an ordered product of integers, each factor greater than one, denoted by $H(n)$. This counting function from number theory was shown by Newberg and Naor (Adv. Appl. Math. 14 (1993) 172–183) to be a lower bound on the number of solutions to the so-called probed partial digest problem, which arises in the analysis of data from experiments in molecular biology. Hille (Acta Arith. 2 (1) (1936) 134–144) established a relation between $H(n)$ and the Riemann zeta function ζ. This relation was used by Hille to prove tight asymptotic upper and lower bounds on $H(n)$. In particular, Hille showed an existential lower bound on $H(n)$: For any $t < \rho = \zeta^{-1}(2) \approx 1.73$ there are infinitely many n which satisfy $H(n) > n^t$. Hille also proved an upper bound on $H(n)$, namely $H(n) = O(n^\rho)$. In this work, we show an improved upper bound on the function $H(n)$, by proving that for every n, $H(n) < n^\rho$ (so 1 can be used as the constant in the 'O' notation). We also present several explicit sequences $\{n_i\}$ with $H(n_i) = \Omega(n_i^d)$, where $d > 1$ is a constant. One sequence has elements of the form $2^\ell 3^j$, and they satisfy $H(n_i) \geqslant n_i^{t_i}$, where $\lim_{i\to\infty} t_i = t \approx 1.43$. This t is the maximum constant for sequences whose elements are products of two distinct primes. Another sequence has elements that are products of four distinct primes, and they satisfy $H(n_i) > n_i^d$, where $d \approx 1.6$.

1. Introduction

Denote by $H(n)$ the number of representations of the natural number n as an ordered product of factors greater than one. Two representations are considered identical if they

* Corresponding address: Massey University, Institute of Fundamental Sciences, Private Bag 11-222, Palmerston North, New Zealand.

E-mail addresses: benny@cs.technion.ac.il (B. Chor), plemke@ccr-p.ida.org (P. Lemke), zivmador@microsoft.com (Z. Mador)

[1] Research supported by the fund for promotion of research at the Technion.

[2] Present address: Microsoft R&D Center, Matam, Haifa 31905, Israel.

0012-365X/00/$ - see front matter
PII: S0012-365X(99)00223-X

contain the same factors in the same order. For example, the different representations of 12 are $12 = 6 \cdot 2 = 2 \cdot 6 = 4 \cdot 3 = 3 \cdot 4 = 3 \cdot 2 \cdot 2 = 2 \cdot 3 \cdot 2 = 2 \cdot 2 \cdot 3$, so $H(12) = 8$.

Our interest in the function $H(n)$ stems from its relation to the analysis of the *probed partial digest problem* (PPDP) in computational biology. Digestion (or restriction) techniques play a central role in molecular biology.[3] A long DNA segment, viewed as a long word over the four-letter alphabet $\{A, C, G, T\}$, is digested by a restriction enzyme. The enzyme identifies the locations where a specific short DNA subsequence occurs, and performs a chemical reaction that cleaves the DNA in those locations. For example, the enzyme EcoRI cuts at the occurrences of *GAATTC*. The lengths (number of letters) of each fragment are then measured. Various digestion techniques give rise to a number of computational problems. We briefly describe three such problems. They all have as input the lengths of fragments whose endpoints are cutting sites on the original DNA segment. The common goal is to identify the locations of these cutting sites (endpoints of the fragments) relative to the ends of the original long DNA segment.

In the double digest problem (DDP) two different restriction enzymes are involved. Each enzyme cuts the DNA at the locations of its particular subsequence. The DNA is completely digested in each of the three ways: By the first enzyme solely, by the second enzyme solely, and by both the enzymes. The problem is to determine the locations of all the cutting sites, given the fragments' lengths from each of the three digestion processes. Goldstein and Waterman [3] proved that the related decision problem (given the fragments' lengths, is there a feasible solution?) is *NP-Complete*. This intractability result implies that there is probably no polynomial time algorithm which solves the DDP decision problem (and thus the search problem). They also discussed the number of solutions an input to the problem can have, and showed that when the restriction sites are modeled by a Poisson process, the number of solutions increases exponentially as the length of the original DNA segment increases. Schmitt and Waterman [9] have further studied and characterized the solutions to DDP.

Partial digest of DNA is another mapping technique. Here, the digestion experiment produces a multiset of the lengths of all the fragments whose endpoints are cutting sites (a multiset is a set whose elements' multiplicities might be more than one). For k cutting sites, the multiset is of size $\binom{k}{2}$. Given this multiset, the partial digest problem (PDP) is once again to determine the locations of the cutting sites. Skiena et al. [10] showed polynomial upper and lower bounds on the number of solutions to an input of the problem.

Different information can be derived by hybridizing a probe to the DNA at some specific location, and measuring only the lengths of fragments which contain the probe. Viewing the original DNA as a sequence of length ℓ, denote by k the unique location of the probe ($1 < k < \ell$). We are given as input the lengths $b - a$ of fragments $[a, b]$ that have cutting sites at both the ends (a and b), and contain the probe inside ($a < k < b$).

[3] Digestion is used, for example, in fingerprinting DNA segments via gel electrophoresis, or in DNA amplification via cloning (see, e.g. [1, pp. 52–60]).

(The probe location k is not part of the input.) The probed partial digest problem (PPDP) is to locate the cutting sites given such data. In all these problems, reversal and additive shift of a solution to an input $\mathcal{M}$ are also solutions to $\mathcal{M}$, and they are all considered congruent to each other. Naor and Newberg [8] proved that the input set of n lengths $\{1,2,\ldots,n\}$ has at least $H(n)$ non-congruent PPDP-solutions. Thus $H(n)$ is a lower bound on the number of PPDP-solutions for an input of size n in the worst case. It should be realized that these bounds are not directly applicable to the real experimental problem, due to noisy inputs.

Hille [5] proved a close relation between $H(n)$ and the Riemann zeta function, $\zeta(t)$. Let ρ be the value of $\zeta^{-1}(2) \cong 1.72864724$. Hille showed that for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{H(n)}{n^{\rho-\varepsilon}} = \infty,$$

implying that for some family of inputs, the number of solutions to PPDP grows faster than $n^{\rho-\varepsilon}$. On the other hand, Hille showed that there exists a universal constant $c > 0$ such that for every n, $H(n) < cn^{\rho}$ (no bounds on c were given).

Despite the above-mentioned lower bound, no explicit sequence $\{n_i\}_{i=1}^{\infty}$ with $H(n_i) = \Omega(n_i^d)$ such that $d > 1$ was known. Newberg and Naor [8] presented an explicit sequence for which $H(n) = \Theta(n \operatorname{polylog} n)$. We demonstrate several explicit sequences $\{n_i\}$ with $H(n_i) = \Omega(n_i^d)$, where for the best sequence $d \simeq 1.605242$. The elements of another sequence are numbers of the form $2^{\ell}3^{j}$, and they satisfy $H(n_i) \geqslant (n_i^{t_i})$, where $t_i \underset{i\to\infty}{\to} t \approx 1.43$. This t is the maximum constant for sequences whose elements are products of two distinct primes. In the other direction, we sharpen Hille's upper bound by showing that for all n, $H(n) < n^{\rho}$.

The remainder of this paper is organized as follows: In Section 2, we give some background on Riemann zeta function. Section 3 describes the improved upper bound on $H(n)$, while Section 4 demonstrates several explicit sequences with fast growing $H(n)$. Finally, Section 5 suggests two open problems.

2. The Riemann zeta function

The Riemann zeta function is defined by $\zeta(t) = \sum_{n\in\mathcal{N}} 1/n^t$ ($\mathcal{N}$ denotes the set of positive integers). The sum converges and the function is well defined for every real number $t > 1$.

Let $\mathcal{B}$ be a finite or infinite set of primes. Let $\mathcal{P}$ be the multiplicative system of all natural numbers which are products (with multiplicity) of primes in $\mathcal{B}$ ($1 \in \mathcal{P}$). The set $\mathcal{B}$ is called the *basis* of the multiplicative system $\mathcal{P}$. For example, the first elements of the multiplicative system over the basis $\{2,3\}$ are $\{1,2,3,4,6,8,9,12,16,\ldots\}$.

The function $\zeta_{\mathcal{P}}(t)$ is defined by summing over $\mathcal{P}$ only: $\zeta_{\mathcal{P}}(t) = \sum_{n\in\mathcal{P}} 1/n^t$. Note that if $\mathcal{B}$ is the set of *all* primes then $\mathcal{P} = \mathcal{N}$ and $\zeta_{\mathcal{P}}$ is the Riemann zeta function. If the basis is infinite then $\zeta_{\mathcal{P}}(t)$ converges for every $t > 1$, and if the basis is finite then $\zeta_{\mathcal{P}}(t)$ converges and is well defined for every positive t (see [4, p. 246]). In its

real convergence range, the function $\zeta_{\mathscr{P}}$ is monotonically decreasing from ∞ to 1, and satisfies:

Proposition 1. $\zeta_{\mathscr{P}}(t) = \prod_{p\in\mathscr{P}} 1/(1-p^{-t})$.

The proof for $\mathscr{P}=\mathscr{N}$ appears in [4], and can be easily generalized for every system $\mathscr{P}$. Let $\rho(\mathscr{P}) \stackrel{\text{def}}{=} \zeta_{\mathscr{P}}^{-1}(2)$, and in particular $\rho \stackrel{\text{def}}{=} \zeta^{-1}(2)$. It is clear that if $\mathscr{P}_1 \subset \mathscr{P}_2 \subset \mathscr{N}$, then for every t, it holds $\zeta_{\mathscr{P}_1}(t) < \zeta_{\mathscr{P}_2}(t) < \zeta(t)$, and therefore $\rho(\mathscr{P}_1) < \rho(\mathscr{P}_2) < \rho$.

Hille proved that for each system $\mathscr{P}$ there exists an absolute constant $c_{\mathscr{P}}$ such that every $n \in \mathscr{P}$ satisfies $H(n) < c_{\mathscr{P}} n^{\rho(\mathscr{P})}$. However, this constant is not given explicitly. In Theorem 5 we show that for every system $\mathscr{P}$, the constant $c_{\mathscr{P}}$ can be taken as 1.

3. Upper bound on $H(n)$

In order to prove Theorem 5 we start with two lemmata:

Lemma 2. *For every multiplicative system $\mathscr{P}$ and for every t in the convergence range of $\zeta_{\mathscr{P}}$,*

$$\sum_{n\in\mathscr{P},n>1} \frac{H(n)}{n^t} = \sum_{\ell\geqslant 1} (\zeta_{\mathscr{P}}(t)-1)^{\ell}.$$

Proof. Consider, for any $n>1$ in $\mathscr{P}$, the coefficient of $1/n^t$ in the expansion of $(\zeta_{\mathscr{P}}(t)-1)^{\ell}$ (see [11, p. 53]). By the definition of $\zeta_{\mathscr{P}}(t)$, it holds $\zeta_{\mathscr{P}}(t)-1 = \sum_{n\in\mathscr{P},n>1} 1/n^t$. Since $\mathscr{P}$ is a multiplicative system, all the divisors of $n \in \mathscr{P}$ are also in $\mathscr{P}$. Therefore the coefficient of $1/n^t$ in $(\sum_{n\in\mathscr{P},n>1} 1/n^t)^{\ell}$ is exactly the number of ordered factorizations of n into ℓ ordered factors in $\mathscr{P}$, each greater than one. Summing over all $\ell \geqslant 1$ yields the number of ordered factorizations of n, namely $H(n)$. □

Lemma 3. *For every $n,i>1$, $H(n^i) > (H(n))^i$.*

Proof. By concatenating any i ordered factorizations of n, we get an ordered factorization of n^i, and therefore $H(n^i) \geqslant (H(n))^i$. The factorization of n^i to a single factor is not an ordered product of i terms, and therefore does not contribute to the right hand side but does contribute to $H(n^i)$. Thus $H(n^i) \geqslant (H(n))^i + 1$. □

Define $t(n) \stackrel{\text{def}}{=} (\log H(n))/(\log n)$ for $n>1$. This function is convenient in 'measuring the exponent' of n in $H(n)$. The following is an immediate corollary of the last lemma:

Corollary 4. *For every $n,i>1$, $t(n^i) > t(n)$.*

We state now the main theorem. We show that the constant $c_{\mathscr{P}}$ can be taken as 1 for every system $\mathscr{P}$, namely,

Theorem 5. *For every multiplicative system $\mathscr{P}$, every $n \in \mathscr{P}$ satisfies*

$$H(n) < n^{\rho(\mathscr{P})}.$$

Proof. Let $t > \rho(\mathscr{P})$. Since $\zeta_{\mathscr{P}}$ is monotonically decreasing, $1 < \zeta_{\mathscr{P}}(t) < 2$. Therefore $\sum_{\ell \geqslant 1} (\zeta_{\mathscr{P}}(t)-1)^{\ell}$ is a converging geometric series, and it equals $(\zeta_{\mathscr{P}}(t)-1)/(2-\zeta_{\mathscr{P}}(t))$. By Lemma 2, we get $\sum_{n\in\mathscr{P}} H(n)/n^t = H(1)+\sum_{\ell\geqslant 1}(\zeta_{\mathscr{P}}(t)-1)^{\ell} = H(1)+1/(2-\zeta_{\mathscr{P}}(t))-1 < \infty$. So $\sum_{n\in\mathscr{P}} H(n)/n^t$ converges, and thus, for $n \in \mathscr{P}$, $H(n) = o(n^t)$.

Assume now, towards a contradiction, that there exists some $n_0 \in \mathscr{P}$ $(n_0 > 1)$ such that $H(n_0) \geqslant n_0^{\rho(\mathscr{P})}$. By Corollary 4, $t_0 \stackrel{\text{def}}{=} t(n_0^2) > t(n_0) \geqslant \rho(\mathscr{P})$. Let us look at the sequence $\{n_0^{2i}\}_{i=2}^{\infty}$. Using Lemma 3, we have $H(n_0^{2i}) > (H(n_0^2))^i = (n_0^{2t_0})^i = (n_0^{2i})^{t_0}$. Therefore, for $t_0 > \rho(\mathscr{P})$ there exists infinitely many n's in $\mathscr{P}$ for which $H(n) > n^{t_0}$, contradicting $H(n) = o(n^{t_0})$. □

Note that the proof cannot be extended for constant $c_{\mathscr{P}} < 1$, since an assumption $H(n_0) \geqslant c_{\mathscr{P}} n_0^{\rho(\mathscr{P})}$ does not imply $t(n_0) \geqslant \rho(\mathscr{P})$. The last theorem implies:

Corollary 6. *For every multiplicative system $\mathscr{P}$, every $n \in \mathscr{P}$ satisfies*

$$t(n) < \rho(\mathscr{P}).$$

4. Explicit lower bounds on $H(n)$

Hille argued that for any $\varepsilon > 0$ there are infinitely many values of n for which $H(n) > n^{\rho-\varepsilon}$, or equivalently, $t(n) > \rho - \varepsilon$ (a detailed proof appears in [8]). This lower bound can be generalized for a multiplicative system $\mathscr{P}$ over any basis, namely for any $\varepsilon > 0$ there are infinitely many values of $n \in \mathscr{P}$ with $t(n) > \rho(\mathscr{P}) - \varepsilon$. Yet, no explicit sequences with $\lim_{n\to\infty} t(n) > 1$ were known. An explicit sequence which satisfies $H(n) = \Theta(n \operatorname{polylog} n)$ is presented in [8]. That sequence satisfies $\lim_{n\to\infty} t(n) = 1$. We present several explicit sequences with t-limits greater than 1.

First we notice that Corollary 4 implies that every n with $t(n) > 1$ gives rise to a sequence $\{n^i\}_{i=2}^{\infty}$ in which $t(n^i) > t(n) > 1$. Such an example is $n = 216$: For this number $H(216) = 252$, so $t(216) \simeq 1.03$. To get a sequence of monotonically increasing t values, one can take $\{n^{2^i}\}_{i=1}^{\infty}$. Every term is the square of its predecessor in the sequence, so by Corollary 4 has a greater t value.

We first look at systems over bases of size two. The members of such system are all the products of powers of two specific primes. The function $H(n)$ depends only on the multiplicities of the primes which compose n, and not on the primes themselves. For example, $H(2^i 3^j) = H(5^i 13^j)$. So, it is clear that among all these systems, we will find the sequence with the greatest t values in the system $\mathscr{P}$ over the basis $\mathscr{B} = \{2, 3\}$.

Note that indeed $\rho(\mathscr{P})$ is maximal in comparison with any other system over a basis of size two. By Proposition 1, $\rho(\mathscr{P})$ is the root of the equation:

$$\frac{1}{1-2^{-t}}\frac{1}{1-3^{-t}}=2,$$

so $\rho(\mathscr{P})\simeq 1.435279084$.

We define $H(1)=\frac{1}{2}$, a definition which is justified in the following expressions for $H(n)$. Hille used Dedekind's inversion formula to find a recursive rule for $H(n)$ [5]:

Proposition 7. *Let $p_1,\ldots,p_k$ be all the distinct primes which divide a natural number n. Then*

$$H(n)=2\left(\sum_{p_i}H\left(\frac{n}{p_i}\right)-\sum_{p_i,p_j}H\left(\frac{n}{p_ip_j}\right)+\cdots+(-1)^{k-1}H\left(\frac{n}{p_1\ldots p_k}\right)\right).$$

Proposition 7 implies that for a prime p, $H(p^i)=2^{i-1}, i\geqslant 0$ (see also [8]), and for products of two primes,

$$H(p^iq^j)=\begin{cases}2^{j-1} & \text{if } i=0,\\ 2^{i-1} & \text{if } j=0,\\ 2(H(p^{i-1}q^j)+H(p^iq^{j-1})-H(p^{i-1}q^{j-1})) & \text{otherwise.}\end{cases}$$

We have looked for a sequence in $\mathscr{P}$ whose t values approach $\rho(\mathscr{P})$. We apply generating functions in order to derive a combinatorial expression for $H(p^iq^j)$. We use this expression to maximize $t(p^iq^j)$, for $i=\lambda\cdot j$, where $\lambda\geqslant 1$ is some constant. Empirical tests show that for a constant ratio $\lambda=i/j$, the values $\{t(p^{\lambda\cdot j}q^j)\}_{j=1}^{\infty}$ tend to some limit, and we would like to maximize its value. In the sequel, we describe a way to achieve the limit $\rho(\mathscr{P})$.

Proposition 8. *For distinct primes p,q, and natural powers $i\geqslant j$,*

$$H(p^iq^j)=2^{i+j-1}\sum_{k=0}^{j}\binom{i}{k}\binom{j}{k}2^{-k}.$$

Proof. Define the ordinary generating function (see [6, p. 81])

$$F_j(x)\stackrel{\text{def}}{=}\sum_{i=0}^{\infty}H(p^iq^j)x^i.$$

First we find an expression for $F_j(x)$, using the recursive rule of $H(p^iq^j)$. For $j=0$, $H(p^i)=2^{i-1}$, so we have

$$F_0(x)=\sum_{i=0}^{\infty}2^{i-1}x^i=\frac{1}{2(1-2x)}.$$

For $j > 0$, by the recursive rule

$$\sum_{i=1}^{\infty} H(p^i q^j)x^i = 2\sum_{i=1}^{\infty} H(p^{i-1}q^j)x^i + 2\sum_{i=1}^{\infty} H(p^i q^{j-1})x^i - 2\sum_{i=1}^{\infty} H(p^{i-1}q^{j-1})x^i.$$

So,

$$F_j(x) - 2^{j-1} = 2xF_j(x) + 2(F_{j-1}(x) - 2^{j-2}) - 2xF_{j-1}(x).$$

This implies that for $j > 0$,

$$F_j(x) = \frac{2(1-x)}{1-2x}F_{j-1}(x) = \left(\frac{2(1-x)}{1-2x}\right)^j F_0(x) = \frac{2^{j-1}}{(1-2x)}\left(1 + \frac{x}{1-2x}\right)^j.$$

By expanding $(1+y)^j$ and then expanding $1/(1-z)^{k+1}$ to a power series, we get

$$F_j(x) = 2^{j-1}\sum_{k=0}^{j}\binom{j}{k}x^k\frac{1}{(1-2x)^{k+1}} = 2^{j-1}\sum_{k=0}^{j}\binom{j}{k}x^k\sum_{\ell=0}^{\infty}\binom{\ell+k}{k}2^\ell x^\ell.$$

By the definition of $F_j(x)$, $H(p^i q^j)$ is the coefficient of x^i in the power series expansion of $F_j(x)$. The term x^i appears in the right-hand side of the last equation whenever $\ell = i - k$. Summation over these indices yields the equality:

$$H(p^i q^j) = 2^{j-1}\sum_{k=0}^{j}\binom{j}{k}\binom{i}{k}2^{i-k} = 2^{i+j-1}\sum_{k=0}^{j}\binom{i}{k}\binom{j}{k}2^{-k}. \qquad \square$$

We will look at values of $n = p^i q^j$ which 'lie on the line' $i = \lambda j$ for some constant $\lambda \geqslant 1$. The last proposition implies

$$H(p^{\lambda j}q^j) = \sum_{k=0}^{j}\binom{\lambda j}{k}\binom{j}{k}2^{(\lambda+1)j-k-1}. \tag{1}$$

Denote by Ent the binary entropy function

$$\mathrm{Ent}(p) = -p\log_2(p) - (1-p)\log_2(1-p)$$

for $0 < p < 1$. Using Stirling formula, it can be shown [2, p. 530]

$$\binom{j}{k} \geqslant \frac{2^{j\mathrm{Ent}(k/j)}}{\sqrt{8k(1-k/j)}}.$$

So it follows that the kth term of expression (1) satisfies

$$\binom{\lambda j}{k}\binom{j}{k}2^{(\lambda+1)j-k-1} \geqslant \frac{2^{\lambda j\mathrm{Ent}(k/\lambda j)}}{\sqrt{8k(1-k/\lambda j)}}\frac{2^{j\mathrm{Ent}(k/j)}}{\sqrt{8k(1-k/j)}}\frac{2^{(\lambda+1)j-k}}{2}$$
$$= \frac{2^{j(\lambda\mathrm{Ent}(k/\lambda j)+\mathrm{Ent}(k/j)+\lambda+1-k/j)}}{16k\sqrt{(1-k/\lambda j)(1-k/j)}}.$$

For every $0\leqslant k\leqslant j$, the last expression is a lower bound to $H(p^{\lambda j}q^{j})$. Therefore, if we look at the system $\mathscr{P}$ over $\{2,3\}$, we have that whenever λj is integer

$$t(2^{\lambda j}3^{j}) = \frac{\log H(2^{\lambda j}3^{j})}{\log(2^{\lambda j}3^{j})}$$
$$> \frac{j(\lambda\mathrm{Ent}(k/\lambda j)+\mathrm{Ent}(k/j)+\lambda+1-(k/j))}{j\log(2^{\lambda}3)}$$
$$-\frac{\log(16k\sqrt{(1-k/\lambda j)(1-k/j)})}{j\log(2^{\lambda}3)} \qquad (2)$$

(all logarithms are to base 2).

Since $0\leqslant k\leqslant j$, we have

$$\frac{\log(16k\sqrt{(1-k/\lambda j)(1-k/j)})}{j\log(2^{\lambda}3)} < \frac{\log(16j)}{j\log(2^{\lambda}3)}.$$

Denote the ratio k/j by r. When j tends to infinity, the right hand side of the last inequality tends to 0. Therefore when j tends to infinity with r held fixed, expression (2) becomes

$$\frac{\lambda\mathrm{Ent}(\frac{r}{\lambda})+\mathrm{Ent}(r)+\lambda+1-r}{\log(2^{\lambda}3)}.$$

Denote the last expression by $C(\lambda,r)$. To maximize it, we first take the derivative with respect to r (we switch from $\log_2$ to ln in order to simplify the derivatives)

$$\frac{\partial C(\lambda,r)}{\partial r} = \frac{\ln(1-r/\lambda)+\ln(\lambda/r)-\ln(r)+\ln(1-r)-\ln(2)}{\ln(3)+\lambda\ln(2)}.$$

Equating the derivative to 0, we get $\lambda=(r(r+1))/(1-r)$ at a local maximum. Denote by $D(r)$ the result of substituting this value of λ back in $C(\lambda,r)$. After simplification we get

$$D(r) = \frac{r(r+1)\ln(1+1/r)+(1-r)\ln(2/(1-r))}{r(r+1)\ln(2)+(1-r)\ln(3)}.$$

Denote the numerator of $D(r)$ by $f(r)$ and its denominator by $g(r)$. To maximize $D(r)$, we look for values of r where the derivative equals 0. This occurs at values of r satisfying $f(r)g'(r)=g(r)f'(r)$.

$$f(r)g'(r) = \left[r(r+1)\ln\left(1+\frac{1}{r}\right)+(1-r)\ln\left(\frac{2}{1-r}\right)\right]$$
$$\times[(2r+1)\ln(2)-\ln(3)],$$

$$g(r)f'(r) = [r(r+1)\ln(2) + (1-r)\ln(3)] \times \left[(2r+1)\ln\left(1+\frac{1}{r}\right) - \ln\left(\frac{2}{1-r}\right)\right].$$

After simplification and organization this gives

$$(1+2r-r^2)\left[\ln(2)\ln\left(\frac{2}{1-r}\right) - \ln(3)\ln\left(1+\frac{1}{r}\right)\right] = 0.$$

Since $1+2r-r^2 > 0$ for $0 < r < 1$, we get

$$\ln(2)\ln\left(\frac{2}{1-r}\right) = \ln(3)\ln\left(1+\frac{1}{r}\right). \tag{3}$$

It is clear that indeed this equation has a solution in the range $0 < r < 1$. Let $r_{\max}$ denote this solution. Using numerical methods, we found that its value is approximately $r_{\max} \simeq 0.586735749$. Let $t_{\max} \stackrel{\text{def}}{=} D(r_{\max})$, then by using the equation $f(r_{\max})g'(r_{\max}) = g(r_{\max})f'(r_{\max})$ we get

$$t_{\max} = \frac{f(r_{\max})}{g(r_{\max})} = \frac{f'(r_{\max})}{g'(r_{\max})} = \frac{(2r_{\max}+1)\ln(1+1/r_{\max}) - \ln(2/(1-r_{\max}))}{(2r_{\max}+1)\ln(2) - \ln(3)}. \tag{4}$$

By substituting (3) in (4) we get

$$t_{\max} = \frac{\ln(1+1/r_{\max})}{\ln(2)} = \frac{\ln(2/(1-r_{\max}))}{\ln(3)}.$$

Therefore

$$\frac{1}{1-2^{-t_{\max}}}\frac{1}{1-3^{-t_{\max}}} = \frac{1}{1-2^{-\ln(1+1/r_{\max})/\ln(2)}}\frac{1}{1-3^{-\ln(2/1-r_{\max})/\ln(3)}} = 2$$

so $t_{\max} = \rho(\mathscr{P})$. Denote $\lambda_{\max} \stackrel{\text{def}}{=} (r_{\max}(r_{\max}+1))/(1-r_{\max}) \simeq 2.25278278$, $\lambda_j \stackrel{\text{def}}{=} \lfloor \lambda_{\max} j \rfloor / j$, and $k_j \stackrel{\text{def}}{=} \lfloor r_{\max} j \rfloor$. Clearly $\lambda_j \underset{j\to\infty}{\to} \lambda_{\max}$ and $k_j/j \underset{j\to\infty}{\to} r_{\max}$.

We look now at the sequence defined by $n_j = 2^{\lambda_j j}3^j$. The first elements in the sequence are: $\{12, 144, 1728, 41472, 497664, \ldots\}$.

Theorem 9. *The sequence* $\{n_j = 2^{\lambda_j j} \cdot 3^j\}_{j=1}^{\infty}$ *satisfies,*

$$\lim_{j\to\infty} t(n_j) = \rho(\mathscr{P}) \simeq 1.435279084.$$

Proof. For every n_j ($j \geqslant 2$ such that $k_j > 0$) in the sequence we have

$$t(n_j) > C\left(\lambda_j, \frac{k_j}{j}\right) - \frac{\log(16j)}{j\log(2^{\lambda_j}3)}.$$

Therefore,

$$\lim_{j\to\infty} t(n_j) \geqslant \lim_{j\to\infty} C\left(\lambda_j, \frac{k_j}{j}\right) = C(\lambda_{\max}, r_{\max}) = D(r_{\max}) = t_{\max} = \rho(\mathscr{P}).$$

The sequence $\{n_j\}_{j=1}^{\infty}$ is contained in the system $\mathscr{P}$ over the basis $\{2,3\}$. By Corollary 6, $\lim_{j\to\infty} t(n_j) \leqslant \rho(\mathscr{P})$. Since $t_{\max} = \rho(\mathscr{P})$, the limit of the sequence above is optimal in $\mathscr{P}$, and in all the systems over a basis of size two. □

We now turn to explicit sequences from systems over bases of size three. For $n = p^i q^j r^k$ $(i,j,k > 0)$ with three distinct prime divisors p,q,r, Hille's recursive rule is

$$\begin{aligned} H(p^i q^j r^k) = 2(&H(p^{i-1} q^j r^k) + H(p^i q^{j-1} r^k) + H(p^i q^j r^{k-1}) \\ &- H(p^{i-1} q^{j-1} r^k) - H(p^{i-1} q^j r^{k-1}) - H(p^i q^{j-1} r^{k-1}) \\ &+ H(p^{i-1} q^{j-1} r^{k-1})). \end{aligned}$$

We applied the generating functions technique to the case of bases with three distinct primes. It yielded the following expression for $H(p^i q^j r^k)$ (we omit the proof, which is similar to the proof of Proposition 8):

Proposition 10. *For distinct primes p,q,r, and natural powers $i \geqslant j \geqslant k$,*

$$H(p^i q^j r^k) = \sum_{\ell=0}^{j} (-1)^\ell \binom{j}{\ell} \binom{i+j-\ell}{j} H(p^{i+j-\ell} r^k).$$

The main drawback of the last expression is that the sum has alternating signs. So it cannot be lower bounded by one of its terms (as we did in expression (2)), and we were not able to use it in order to find an explicit sequence composed of three primes.

However, dynamic programming is helpful in calculating the value $H(p^i q^j r^k)$. This is done in stages, where in stage a $(0 \leqslant a \leqslant i)$ we calculate the values of $H(p^a q^b r^c)$ with $0 \leqslant b \leqslant j$ and $0 \leqslant c \leqslant k$ in the increasing order of b and c, and store them. In the next stage, all the values needed to calculate $H(p^{a+1} q^b r^k)$ are already known and stored. We actually use the values of only the last two stages, so only these values should be kept in memory. Thus the space required for computing $H(p^i q^j r^k)$ is $\Theta(jk)$. The values which were computed are $H(p^a q^b r^c)$ with $0 \leqslant a \leqslant i$, $0 \leqslant b \leqslant j$, $0 \leqslant c \leqslant k$, and their number is $\Theta(ijk)$, even if we compute only those cases where $a \geqslant b \geqslant c$. Each value is computed exactly once, so calculation of $H(p^i q^j r^k)$ takes time $\Theta(ijk)$ (further details on this computation can be found in [7, pp. 55–60]). As before, it is worthwhile to look at the basis of the first three primes $\{2,3,5\}$. For $\mathscr{P}$ over this basis, $\rho(\mathscr{P}) \simeq 1.56603$. We searched for $n \in \mathscr{P}$ with large t value. Then we went on to look for integers over the basis $\{2,3,5,7\}$ (using the appropriate version of Hille's recursive rule). We found the following sequences:

Theorem 11. *The powers sequence of $n_3 = 2^{1020} \cdot 3^{441} \cdot 5^{177}$, which belongs to the system over $\{2,3,5\}$, satisfies*

$$H(n) > n^{1.56065}.$$

The powers sequence of $n_4 = 2^{263} \cdot 3^{106} \cdot 5^{43} \cdot 7^{24}$, which belongs to the system over $\{2,3,5,7\}$, satisfies

$$H(n) > n^{1.60524}.$$

Proof. Use Corollary 4 with the specified n_3, for which $t(n_3) \simeq 1.56065$, and with n_4, for which $t(n_4) \simeq 1.60524$. □

When further enlarging the base size, the time and space complexities of the computations increase substantially, while the $\rho(\mathscr{P})$ value becomes only slightly larger. For example, the system $\mathscr{P}$ over $\mathscr{B} = \{2,3,5,7\}$ satisfies $\rho(\mathscr{P}) \simeq 1.62705$, while the system $\mathscr{P}$ over $\mathscr{B} = \{2,3,5,7,11\}$ satisfies $\rho(\mathscr{P}) \simeq 1.65257$. Recall that for $\mathscr{P}$ over all the primes $\rho(\mathscr{P}) = \rho \simeq 1.72864$.

5. Open problems

(i) Find explicit sequences whose t-limits equal $\rho(\mathscr{P})$ for systems $\mathscr{P}$ over larger bases, for example $\{2,3,5\}$ and $\{2,3,5,7\}$.

(ii) Find an explicit sequence with t-limit which equals $\rho = \rho(\mathscr{N})$. Such a sequence would be optimal over all $\mathscr{N}$. Note that such an optimal sequence cannot be included in any system over a finite basis.

Acknowledgements

We would like to thank Dalit Naor for her helpful suggestions and remarks.

References

[1] N.G. Cooper (Ed.), The Human Genome Project, University Science books, Mill Valley, CA, 1994.
[2] R. Gallager, Information Theory and Reliable Communication, Wiley, New York, 1968.
[3] L. Goldstein, M.S. Waterman, Mapping DNA by stochastic relaxation, Adv. App. Math. 8 (1987) 194–207.
[4] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, 5th Edition, Oxford, 1985.
[5] E. Hille, A problem in factorisatio numerorum, Acta Arith. 2 (1) (1936) 134–144.
[6] C.L. Liu, Introduction to Combinatorial Mathematics, McGraw-Hill, New York, 1968.
[7] Z. Mador, The probed partial digest problem, algorithms and number of solutions, Master's Thesis, Dept. of Computer Science, Technion, IIT, 1996 (Written in Hebrew).
[8] L.A. Newberg, D. Naor, A lower bound on the number of solutions to the probed partial digest problem, Adv. Appl. Math. 14 (1993) 172–183.
[9] W. Schmitt, M.S. Waterman, Multiple solutions of DNA restriction mapping problems, Adv. Appl. Math. 12 (1991) 412–427.
[10] S.S. Skiena, W.D. Smith, P. Lemke, Reconstructing sets from interpoint distances, Proceedings of 5th Annual Symposium on Computational Geometry, 1990, pp. 332–339.
[11] H.S. Wilf, Generatingfunctionology, Academic Press, New York, 1990.