Longest common extensions

Béla Bollobás [*†‡] Shoham Letzter [*]

June 12, 2016

Abstract

Given a word w of length n and $i, j \in [n]$, the *longest common extension* is the longest substring starting at both i and j. In this note we estimate the average length of the longest common extension over all words w and all pairs (i, j), as well as the typical maximum length of the longest common extension.

We also consider a variant of this problem, due to Blanchet-Sadri and Lazarow, in which the word is allowed to contain 'holes', which are special symbols functioning as 'jokers', i.e. are considered to be equal to any character. In particular, we estimate the average longest common extension over all words w with a small number of holes, extending a result by Blanchet-Sadri, Harred and Lazarow, and prove a similar result for words with holes appearing randomly.

1 Introduction

Given a word w of length n and $i, j \in [n]$, the *longest common extension* of i and j is the longest substring which starts at both i and j. We denote the length of this substring by $l_w(i, j)$. In other words, $l_w(i, j)$ is the maximum l such that $w_{i+t} = w_{j+t}$ for every $0 \leq t < l$. The problem of finding the maximum longest common extension among a list of given pairs appears as a subproblem of many known problems regarding substrings, such as the k-mismatch problem and the k-difference global alignment problem (see [12, 13, 15, 1, 10, 2] and [16, 5] for more recent developments), estimating the number of tandem repeats (see [8, 11, 14]), and computing palindromes or matchings with holes (see [7]).

We denote by $f(n, k)$ the average of $l_w(i, j)$ over all words w of length n over alphabet $[k] = \{1, \ldots, k\}$ and all pairs (i, j) with $1 \leq i < j \leq n$. Ilie, Navarro and Tinta [9] computed $f(n, k)$ exactly and obtained the following result.

[*]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, CB3 0WB Cambridge, UK; e-mail: {B.Bollobas,S.Letzter}@dpmms.cam.ac.uk.

[†]Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA and London Institute for Mathematical Sciences, 35a South St., Mayfair, London W1K 2XF, UK.

[‡]Research supported in part by NSF grant DMS-1301614 and EU MULTIPLEX grant 317532.

Theorem 1 (Ilie, Navarro and Tinta [9]). *Let $k \geq 2$ be fixed. Then $\lim_{n\to\infty} f(n,k) = 1/(k-1)$.*

A modification of this problem has been introduced and studied by Blanchet-Sadri and Lazarow [4], with further results proved by Crochemore et al. [6] and Blanchet-Sadri, Harred and Lazarow [3]. Given alphabet A, let w be a word over alphabet $A \cup \{\diamondsuit\}$, where $\diamondsuit$ stands for a 'hole'. We let $l_w(i,j)$ be the largest l such that for every $0 \leq t < l$ either w_{i+t} and w_{j+t} are equal or one of them is a hole.

Our first aim in this paper is to consider, as in [3], the analogue $f(n,k,h)$ of $f(n,k)$ for words with h holes. In other words, $f(n,k,h)$ is the average of $l_w(i,j)$ over all words w of length n over alphabet $[k] \cup \{\diamondsuit\}$ with exactly h $\diamondsuit$'s, and over all pairs (i,j) with $1 \leq i < j \leq n$. We shall show that, perhaps unsurprisingly, when the number of holes is small, the effect on the average longest common extension is negligible. This extends the result of Blanchet-Sadri, Harred and Lazarow [3] who showed that if h is constant, then $\lim_{n\to\infty} f(n,k,h) = \frac{1}{k-1}$. The following result implies that this holds whenever $h = o(n^{1/3})$.

Theorem 2. *Let $k = k(n) \geq 2$ and $n \geq h$. Then, as $n \to \infty$, $f(n,k,h) = \frac{1}{k-1} + O\left(\frac{h^3}{n}\right)$.*

Our next aim is to consider the variant of $f(n,k,h)$ in which the holes appear randomly. For $p \in [0,1]$ denote by $g(n,k,p)$ the average of $l_w(i,j)$ where w is an n-letter word each of whose letters is chosen independently to be a hole with probability p and any letter in $[k]$ with probability $\frac{1-p}{k}$, and over all pairs (i,j) with $1 \leq i < j \leq n$. This model is very close to the case of random words of length n over an alphabet of size k, containing approximately pn holes.

Theorem 3. *Let $k = k(n) \geq 2$ and $p = p(n) \in [0,1)$. Then, as $n \to \infty$, $g(n,k,p) = \frac{q}{1-q} + O\left(\frac{q}{(1-\sqrt{q})^2 n}\right)$, where $q = 1 - \frac{(1-p)^2(k-1)}{k}$.*

To better understand the estimate of $g(n,k,p)$ from Theorem 3, note that

$$\begin{aligned}\frac{q}{1-q} &= \frac{1 - \frac{(1-p)^2(k-1)}{k}}{\frac{(1-p)^2(k-1)}{k}} \\ &= \frac{1}{k-1} \cdot \left(\frac{k}{(1-p)^2} - (k-1)\right) \\ &= \frac{1}{k-1} \cdot \left(1 + k \cdot \left(\frac{1}{(1-p)^2} - 1\right)\right).\end{aligned}$$

In particular, if $p = 0$, the estimate in Theorem 3 coincides (as expected) with the estimate of $f(n,k)$. If $p = \frac{1}{k+1}$, i.e., a hole appears with the same probability as the characters from $[k]$, we find that $g(n,k,p) = \frac{3}{k-1} + \frac{1}{k(k-1)} + O\left(\frac{1}{kn}\right)$. Finally, we consider the case where p is close to 1. Set $p = 1 - \varepsilon$, and assume that ε is 'small'. Then $f(n,k,p) = \frac{k}{k-1} \cdot \frac{1}{\varepsilon^2} - 1 + O\left(\frac{1}{\varepsilon^4 n}\right)$ (so this estimate is useful when $\varepsilon = \omega(n^{-1/2})$).

Our final aim in this note is to study the maximum, rather than average, of the length of the longest common extensions. To be precise, as before, let w be a word of length n, and set $l(w) =$

$\max_{i<j} l_w(i,j)$. Theorem 4 below shows that $l(w)$ is, with high probability, very close to $2\log_k n$. This improves a result of Ilie, Navarro and Tinta [9] who showed that the average of $l(w)$ is at least $\log_k n - 2$ and at most $2\log_k n$ (in fact, there is a mistake in the calculation of their upper bound: they use an upper bound on the number of strings of length n and alphabet $[k]$ with a repeated substring of length l, but their argument gives a worse upper bound than stated, which results in an ineffective upper bound for the expectation). We note that statements about events that occur with high probability tend to be more interesting and harder to prove than statements about expectations.

Theorem 4. *Let n and $k = k(n)$ be such that $k \geq 2$ and $\log_k n \to \infty$ as $n \to \infty$. If w is chosen uniformly at random from $[k]^n$, then $l(w) = 2\log_k n + O(\log_k(\log_k n))$, with high probability.*

We prove Theorems 1 to 3 in Section 2 and prove Theorem 4 in Section 3.

2 Average length of longest common extensions

In this section we shall prove Theorems 1 to 3, which concern the average longest common extension in several settings. Before we turn to the proofs, we mention the following equality, which holds for $|x| < 1$. We shall use it several times throughout this section.

$$\sum_{l\geq 0}(l+1)x^l = \sum_{s\geq 0}\sum_{t\geq s} x^t = \left(\frac{1}{1-x}\right)^2. \tag{1}$$

For the sake of completeness, we give a short proof of Theorem 1, due to Ilie, Navarro and Tinta [9]. Recall that $f(n,k)$ is the average of the common longest extensions for words of length n over alphabet $[k]$. In fact, we shall prove Theorem 1 in the following quantitative form.

Theorem 1'. *Let $k \geq 2$. Then $f(n,k) = \left(1 + O\left(\frac{1}{kn}\right)\right)\frac{1}{k-1}$.*

Proof. The following holds, where w is taken uniformly at random from $[k]^n$.

$$\begin{aligned} f(n,k) &= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}_w[l_w(i,j)] \\ &= \frac{1}{\binom{n}{2}}\sum_{l\geq 0}\sum_{i<j}\mathbb{P}[l_w(i,j) > l]. \end{aligned}$$

We now evaluate the probability that $l_w(i,j) > l$ (where $i \neq j$).

$$\mathbb{P}[l(i,j) > l] = \begin{cases} \left(\frac{1}{k}\right)^{l+1} & \text{if } i,j \leq n-l \\ 0 & \text{otherwise} \end{cases}$$

Indeed, w_i is equal to w_j with probability $1/k$, w_{i+1} is equal to w_{j+1} with probability $1/k$ and so on. This reasoning holds even if $i \geq n-l$. In particular, we have that $\mathbb{P}[l(i,j) > l] \leq \left(\frac{1}{k}\right)^{l+1}$. Thus, by the above expression for $f(n,k)$,

$$f(n,k) \leq \sum_{l \geq 0} \left(\frac{1}{k}\right)^{l+1} = \frac{1}{k} \cdot \frac{1}{1-1/k} = \frac{1}{k-1}.$$

We next give a lower bound on $f(n,k)$ which is very close to the upper bound. Here we use (1) and the fact that $\binom{n}{2} - \binom{n-l}{2} = (2nl - l - l^2)/2 \leq nl$.

$$\begin{aligned}
f(n,k) &= \frac{1}{\binom{n}{2}} \sum_{l \geq 0} \binom{n-l}{2} \left(\frac{1}{k}\right)^{l+1} \\
&\geq \frac{1}{\binom{n}{2}} \sum_{l \geq 0} \left(\binom{n}{2} - nl\right) \left(\frac{1}{k}\right)^{l+1} \\
&= \sum_{l \geq 0} \left(\frac{1}{k}\right)^{l+1} - \frac{2}{n-1} \sum_{l \geq 0} l \left(\frac{1}{k}\right)^{l+1} \\
&= \frac{1}{k-1} - \frac{2}{n-1} \left(\frac{1}{k-1}\right)^2 \\
&= \left(1 + O\left(\frac{1}{kn}\right)\right) \frac{1}{k-1}.
\end{aligned}$$

To sum up, we have $f(n,k) = \left(1 + O\left(\frac{1}{kn}\right)\right) \frac{1}{k-1}$. □

We proceed to the proof of Theorem 2 which gives an estimate for $f(n,k,h)$, the average of the common longest extensions of words of length n with h holes over alphabet $[k]$.

Theorem 2. *Let $k = k(n) \geq 2$ and $n \geq h$. Then, as $n \to \infty$, $f(n,k,h) = \frac{1}{k-1} + O\left(\frac{h^3}{n}\right)$.*

Proof. Clearly, the probability $\mathbb{P}[l_w(i,j)]$ can only increase by the appearance of holes. Thus,

$$f(n,k,h) \geq f(n,k) = \left(1 + O\left(\frac{1}{n}\right)\right) \frac{1}{k-1}.$$

We now obtain an upper bound on $f(n,k,h)$. We use the following crude bounds on the probability that $l_w(i,j) > l$ for a word w with at most h holes (here $x^+ = \max\{0,x\}$).

$$\mathbb{P}[l(i,j) > l] \leq \begin{cases} \left(\frac{1}{k}\right)^{l+1} & \text{if none of } w_{i+k}, w_{j+k} \text{ is a hole for } k \leq l \\ \left(\frac{1}{k}\right)^{(l+1-2h)^+} & \text{otherwise} \end{cases}$$

In order to obtain an upper bound, we note that the following inequality holds.

$$f(n,k,h) \leq \frac{1}{\binom{n}{2}} \left(\binom{n}{2} \sum_{l \geq 0} \left(\frac{1}{k}\right)^{l+1} + 8h^3 n + \sum_{l \geq 2h} hln \left(\frac{1}{k}\right)^{l+1-2h} \right).$$

To see this, the first term accounts for all choices of $i < j$ such that $w_{i+t}, w_{j+t} \neq \diamondsuit$ for every $t \leq l$; the second term accounts for all pairs $i < j$ such that $w_{i+t} = \diamondsuit$ or $w_{j+t} = \diamondsuit$ for some $t \leq l$ where $l \leq 2h-1$ (h choices for a particular hole, $2h$ choices for l, $l+1 \leq 2h$ choices for t, two choices for which of w_{i+t} and w_{j+t} is a hole, and n choices for either i or j); and the third term accounts for pairs $i < j$ for which $w_{i+t} = \diamondsuit$ or $w_{j+t} = \diamondsuit$ for some $t < l$ and $l \geq h$.

The contribution of the first term to the sum is exactly $\frac{1}{k-1}$; the contribution of the second term is $O\left(\frac{h^3}{n}\right)$; and to calculate the contribution of the third term, using (1), we have the following.

$$\begin{aligned} \sum_{l \geq 2h} hl \left(\frac{1}{k}\right)^{l+1-2h} &= \sum_{t \geq 0} h(t+2h-1) \left(\frac{1}{k}\right)^{t} \\ &= h\frac{k}{(k-1)^2} + h(2h-1)\frac{k}{k-1} = O(h^2) \end{aligned}$$

It follows that the contribution of the third term to the sum is $O\left(\frac{h^2}{n}\right)$. Consequently,

$$f(n,k,h) = \frac{1}{k-1} + O\left(\frac{h^3}{n}\right).$$

We note that this estimate holds even if the h holes are in prescribed positions in the word (rather than choosing the positions of the holes uniformly at random). □

We now prove Theorem 3 which gives an estimate for $g(n,k,p)$, the average of the longest common extensions in words of length n over alphabet $[k]$ and holes appearing with probability p.

Theorem 3. *Let $k = k(n) \geq 2$ and $p = p(n) \in [0,1)$. Then, as $n \to \infty$, $g(n,k,p) = \frac{q}{1-q} + O\left(\frac{q}{(1-\sqrt{q})^2 n}\right)$, where $q = 1 - \frac{(1-p)^2(k-1)}{k}$.*

Proof. We say that two different characters are *compatible* if they are either equal or one of them is a hole. The probability that two different characters are incompatible (i.e. are not holes and are distinct) is $\frac{(1-p)^2(k-1)}{k}$. Hence $q = 1 - \frac{(1-p)^2(k-1)}{k}$ is the probability that two different characters are compatible. We obtain the following bounds for the probability that $l_w(i,j) > l$.

$$\begin{aligned} &\mathbb{P}[l(i,j) > l] = 0 && \text{if } j > n-l \\ &\mathbb{P}[l(i,j) > l] = q^{l+1} && \text{if } j \leq n-l \text{ and } i \leq j-l-1 \\ &\mathbb{P}[l(i,j) > l] \leq \sqrt{q}^{\,l+1} && \text{otherwise.} \end{aligned}$$

To see why the inequality in the third line holds, suppose that $j = i + t$ where $t \leq l$. Denote $I_s = \{i + st, i + st + 1, \ldots, i + (s+1)t - 1\}$ and let $K = (I_0 \cup I_2 \cup \ldots) \cap \{i, \ldots, i + t\}$. Note that the events $\{w_{i+s}, w_{j+s} \text{ are compatible}\}$ are independent for $s \in K$ and have probability q each. Since $|K| \geq (l+1)/2$, it follows that $\mathbb{P}[l(i,j) > l] \leq q^{(l+1)/2} = \sqrt{q}^{\,l+1}$.

We start with the lower bound. Note that there are at least $(n-l)(n-3l)/2$ pairs (i,j) such that $j \leq n - l$ and $i \leq j - l - 1$. For every such pair the probability that $l_w(i,j) > l$ is q^{l+1}. Thus the following upper bound holds (the last equality follows from (1)).

$$\begin{aligned} g(n,k,p) &\geq \frac{1}{\binom{n}{2}} \sum_{l \geq 0} \frac{1}{2}(n-l)(n-3l-1)q^{\,l+1} \\ &\geq \frac{1}{\binom{n}{2}} \sum_{l \geq 0} \left(\binom{n}{2} - 2ln \right) q^{\,l+1} \\ &= \frac{q}{1-q} - \frac{4}{n-1} \left(\frac{q}{1-q} \right)^2 \end{aligned}$$

For the upper bound, we have the following inequality.

$$\begin{aligned} g(n,k,h) &\leq \frac{1}{\binom{n}{2}} \left(\binom{n}{2} \sum_{l \geq 0} q^{\,l+1} + \sum_{l \geq 0} ln \cdot \sqrt{q}^{\,l+1} \right) \\ &= \frac{q}{1-q} + \frac{2}{n-1} \cdot \frac{q}{(1-\sqrt{q})^2}. \end{aligned}$$

Indeed, the second term takes into account the pairs (i,j) such that $j - l \leq i < j$ and its contribution can be bounded using (1). Combining the upper and lower bounds, we have that

$$g(n,k,p) = \frac{q}{1-q} + O\left(\frac{q}{(1-\sqrt{q})^2 n} \right),$$

completing the proof of Theorem 3. □

3 Maximum longest common extensions of random words

In this section we address the problem of finding the maximum length of the longest common extensions. In particular, we prove Theorem 4.

Theorem 4. *Let n and $k = k(n)$ be such that $k \geq 2$ and $\log_k n \to \infty$ as $n \to \infty$. If w is chosen uniformly at random from $[k]^n$, then $l(w) = 2\log_k n + O(\log_k(\log_k n))$, with high probability.*

Proof. Let w be a word selected uniformly at random from $[k]^n$. We first prove an upper bound on $l(w)$. Suppose that $k^{\,l} = \omega(n^2)$. Recall that the probability that $l_w(i,j) \geq l$ is at most $\left(\frac{1}{k}\right)^l$.

Thus, the probability that there is a pair $i < j$ for which $l_w(i,j) \geq l$ is at most $\binom{n}{2}\left(\frac{1}{k}\right)^l = o(1)$, i.e., $\mathbb{P}[l(w_n) \geq l] = o(1)$ if $k^l = \omega(n^2)$. Taking $l = 2\log_k n + \log_k(\log_k n)$, we have $k^l = n^2 \cdot \log_k n = \omega(n^2)$, implying that $\mathbb{P}[l(w) \geq l] = o(1)$. In other words, $l(w_n) \leq 2\log_k n + \log_k(\log_k n)$, with high probability.

On the other hand, if w is an n-letter word satisfying that $l(w) < l$, then, in particular, the subwords $w^{(t)} = w_{(t-1)l+1} \dots w_{tl}$, where $t = 1, \dots, \lfloor \frac{n}{l} \rfloor$, are distinct. Thus, the probability that $l(w) < l$ is at most the probability that $w^{(1)}, \dots, w^{(T)}$ are distinct, which, if w is chosen uniformly at random out of $[k]^n$, is the probability that a sequence $T = \lfloor \frac{n}{l} \rfloor$ words of length l generated independently contains no two words which are the same. We thus obtain the following upper bound on the probability that $l(w) < l$.

$$\begin{aligned}
\mathbb{P}[l(w) < l] &\leq \frac{k^l(k^l-1)\dots(k^l-(T-1))}{(k^l)^T} \\
&= \prod_{0 \leq t < T} \left(1 - \frac{t}{k^l}\right) \\
&\leq \exp\left(-\sum_{0 \leq t < T} \frac{t}{k^l}\right) \\
&= \exp\left(-\frac{\binom{T}{2}}{k^l}\right).
\end{aligned}$$

Hence, if $k^l = o\left(\left(\frac{n}{l}\right)^2\right)$, $\mathbb{P}[l(w) < l] = o(1)$. Let $l = 2\log_k n - 3\log_k(\log_k n)$. Then $k^l = \frac{n^2}{(\log_k n)^3} = o\left(\left(\frac{n}{l}\right)^2\right)$, implying that $\mathbb{P}[l(w) < l] = o(1)$. In other words, $l(w_n) \geq 2\log_k n - 3\log_k(\log_k n)$, with high probability. □

From this high concentration result, it is but a short step to conclude that the expectation of $l(w)$ is about $2\log_k n$, improving a result from [9].

Corollary 5. $\mathbb{E}[l(w)] = 2\log_k(n) + O(\log_k(\log_k n))$.

Proof. We first prove the lower bound. From the proof of Theorem 4, it follows that if $l = 2\log_k n - 3\log_k(\log_k n)$ then $\mathbb{P}[l(w) \leq l] \leq \exp\left(-\binom{n/l}{2}\frac{1}{k^l}\right) = \exp(-\log_k n/8) \leq \frac{1}{\log_k n}$. It follows that

$$\begin{aligned}
\mathbb{E}[l(w)] &\geq l \cdot \mathbb{P}(l(w) \geq l) \geq \\
&(2\log_k n - 3\log_k(\log_k n))\left(1 - \frac{1}{\log_k n}\right) = \\
&2\log_k n + O(\log_k(\log_k n)).
\end{aligned}$$

For the proof of the upper bound we need to be a little more careful. Let $l_1 = 2\log_k n + 2\log_k \log_k n$ and $l_2 = 4\log_k n$. Then $\mathbb{P}(l(w) \geq l_1) \leq \binom{n}{2}\left(\frac{1}{k}\right)^l \leq (\log_k n)^{-2}$ and $\mathbb{P}(l(w) \geq l_2) \leq n^{-2}$. Hence,

$$\begin{aligned}
\mathbb{E}[l(w)] &\leq l_1 \cdot \mathbb{P}(l(w) < l_1) + l_2 \cdot \mathbb{P}(l_1 \leq l(w) < l_2) + n \cdot \mathbb{P}(l(w) \geq l_2) \\
&\leq l_1 + \frac{l_2}{(\log_k n)^2} + \frac{n}{n^2} \\
&= 2\log_k n + O(\log_k(\log_k n)).
\end{aligned}$$

It follows that $\mathbb{E}[l(w)] = 2\log_k n + O(\log_k(\log_k n))$. □

The addition of holes with probability p, as in Theorem 3, would result in a similar theorem, but with a different constant multiple of $\log_k n$ as the main term.

Acknowledgements

BB is grateful to Zsuzsanna Lipták for her generous hospitality at IWOCA 2015 in Verona, and to Francine Blanchet-Sadri for her lecture at the conference introducing him to the problems discussed above.

References

[1] K. Abrahamson, *Generalized string matching*, SIAM Journal on Computing **16** (1987), no. 6, 1039–1051.

[2] A. Amir, M. Lewenstein, and E. Porat, *Faster algorithms for string matching with k mismatches*, Journal of Algorithms, **50** (2004), no. 2, 257–275.

[3] F. Blanchet-Sadri, R. Harred, and J. Lazarow, *Combinatorial Algorithms: 26th International Workshop, IWOCA 2015, Verona, Italy, October 5-7, 2015, Revised Selected Papers*, ch. Longest Common Extensions in Partial Words, pp. 52–64, Springer International Publishing, Cham, 2016.

[4] F. Blanchet-Sadri and J. Lazarow, *Suffix trees for partial words and the longest common compatible prefix problem*, Language and Automata Theory and Applications (Adrian-Horia Dediu, Carlos Martn-Vide, and Bianca Truthe, eds.), Lecture Notes in Computer Science, vol. 7810, Springer Berlin Heidelberg, 2013, pp. 165–176.

[5] R. Clifford, A. Fontaine, E. Porat, B. Sach, and T. Starikovskaya, *The k-mismatch problem revisited*, Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA, USA), SODA '16, Society for Industrial and Applied Mathematics, 2016, pp. 2039–2052.

[6] M. Crochemore, C. S. Iliopoulos, T. Kociumaka, M. Kubica, A. Langiu, J. Radoszewski, W. Rytter, B. Szreder, and T. Waleń, *A note on the longest common compatible prefix problem for partial words*, preprint, arxiv:1312.2381.

[7] D. Gusfield, *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, Cambridge Univ. Press, New York, NY, USA, 1997.

[8] D. Gusfield and J. Stoye, *Linear time algorithm for finding and representing all tandem repeats in a string*, J. Comput. Syst. Sci. **69** (2004), 525–546.

[9] L. Ilie, G. Navarro, and L. Tinta, *The longest common extension problem revisited and applications to approximate string searching*, J. Discrete Algorithms **8** (2010), 418–428.

[10] S. R. Kosaraju, *Efficient string matching*, (1987), manuscript,.

[11] G. Landau, J.P. Schmidt, and D. Sokol, *An algorithm for approximate tandem repeats*, J. Comput. Biol. **8** (2001), 1–18.

[12] G. Landau and U. Vishkin, *Fast parallel and serial approximate string matching*, J. Algorithms **10** (1989), 157–169, preliminary version in: ACM STOC'86.

[13] G. M. Landau and U. Vishkin, *Efficient string matching with k mismatches*, Theoretical Computer Science **43** (1986), 239–249.

[14] M. Main and R.J. Lorentz, *An $O(n \log n)$ algorithm for finding all repetitions in a string*, J. Algorithms **5** (1984), 422–432.

[15] G. Myers, *An $O(nd)$ difference algorithm and its variations*, Algorithmica **1** (1986), 251–266.

[16] M. Nicolae and S. Rajasekaran, *On string matching with mismatches*, Algorithms **8** (2015), no. 2, 248–270.