Corrigendum to "Parameterized Tractability of the Maximum-Duo Preservation String Mapping Problem"

Stefano Beretta[a,*], Mauro Castelli[b], Riccardo Dondi[c]

[a] *Dipartimento di Informatica, Sistemistica e Comunicazione, Università degli Studi di Milano - Bicocca, Milano - Italia*
[b] *NOVA IMS, Universidade Nova de Lisboa, Lisboa - Portugal*
[c] *Dipartimento di Lettere, Filosofia e Comunicazione, Università degli Studi di Bergamo, Bergamo - Italia*

Abstract

This is a corrigendum for our paper [1], as we have found that the first FPT algorithm for the Maximum-Duo Preservation String Mapping Problem we presented is incorrect. However, we show that, by slightly modifying the color-coding technique on which the algorithm is based, we can fix the error, thus giving a correct FPT algorithm for Maximum-Duo Preservation String Mapping Problem.

Keywords: Computational Biology, Common String Partition, Parameterized Algorithms

1. Introduction

In [1], we proved that the Maximum-Duo Preservation String Mapping Problem is FPT, and, in particular, we presented a FPT algorithm based on the color-coding technique. However, as we will show later in this paper, the algorithm is incorrect. By slightly modifying the coloring on which the algorithm is based, we can fix the error, thus giving a correct FPT algorithm for Maximum-Duo Preservation String Mapping Problem.

First, we present the definitions that will be useful for the FPT algorithm that we will describe in Section 2. Most of the definitions are as in [1], we include them for completeness.

Given a string A, a *duo* is an ordered pair of consecutive elements $(A[i], A[i+1])$. Two strings A and B are *related*, if and only if B is a permutation of A. In the rest of the paper we will denote by n the length of A and B. Given two related strings A and B, a *mapping* m of A into B is a bijective function from the positions of A to the positions of B such that $m(i) = j$ implies that $A[i] = B[j]$. A *partial mapping* m of A into B is a bijective function from a subset of positions of A to a subset of positions of B such that $m(i) = j$ implies that $A[i] = B[j]$.

The definition of mapping and partial mapping can be extended to two sets of duos of related strings A and B, that is if positions i and $i+1$, with $1 \leq i \leq n-1$, are mapped into positions j and $j+1$, with $1 \leq j \leq n-1$, we say that duo $(A[i], A[i+1])$ is mapped into duo $(B[j], B[j+1])$.

Given two related strings A and B, and a mapping m of the positions of A into the positions of B, a duo $(A[i], A[i+1])$ is preserved if $m(i) = j$ and $m(i+1) = j+1$.

Now, we give the definition of the Maximum-Duo Preservation String Mapping Problem (in its decision version).

Maximum-Duo Preservation String Mapping Problem (Max-Duo PSM)
Input: two related strings A and B, an integer k.
Output: is there a mapping m of A into B such that the number of preserved duos is at least k?

*Corresponding author
Email addresses: stefano.beretta@disco.unimib.it (Stefano Beretta), mcastelli@novaims.unl.pt (Mauro Castelli), riccardo.dondi@unibg.it (Riccardo Dondi)

Given two positions $1 \leq i < j \leq n$, we denote by $d_{S(i,j)}$ the sequence of *consecutive duos* $(S[i], S[i+1])$, $\ldots$, $(S[j-1], S[j])$. We say that two positions $A[i]$ and $A[i+1]$ *generate* the duo $(A[i], A[i+1])$. Notice that k preserved duos are generated by at most $2k$ positions. We say that position i, with $1 \leq i \leq n-1$, of a string S, induces duo $(S[i], S[i+1])$.

2. An FPT Algorithm for Max-Duo PSM

Here, we briefly discuss the FPT algorithm for Max-Duo PSM we presented in [1] and why it is not correct.

The FPT algorithm is based on the color-coding technique, that assigns a color to each position of the string B, such that to each position i that induces a preserved duo $(B[i], B[i+1])$ is assigned a distinct color. We will show in the following example that such a coloring does not lead to a correct algorithm. Let

$$A = abbc \quad B = abcb$$

be two related input strings. Let $C = \{c_1, c_2\}$ be a set of colors and assume that the coloring assigns to positions 1, 3 and 4 color c_1, and to position 2 color c_2. The dynamic programming we presented in [1], allows us to map duo $(A[3], A[4])$ (that is bc) to duo $(B[2], B[3])$, and duo $(A[1], A[2])$ (that is ab) to duo $(B[1], B[2])$, since positions 1 and 2 in B are associated to distinct colors. Hence, the algorithm in [1] outputs that there exists a solution to Max-Duo PSM when $k = 2$. However, $(B[1], B[2])$ and $(B[2], B[3])$ cannot be both preserved by a solution of Max-Duo PSM, as position 2 of B can be mapped either to position 2 of A, thus allowing to preserve $(B[1], B[2])$, or to position 3 of A, thus allowing to preserve $(B[2], B[3])$. Indeed, an optimal solution of Max-Duo PSM on instance A and B preserves one duo.

Next, we show how to correct the FPT algorithm. The main idea is to assign two distinct colors to positions j and $j+1$ when $(B[j], B[j+1])$ is preserved, thus applying the coloring to positions that generate a duo, instead of positions that induce a duo. Now, we present the details of the algorithm.

Given an integer $t \leq 2k$, let $C = \{c_1, \ldots, c_t\}$ be a set of t *colors*, the algorithm computes whether there exists a set of t positions that generate k duos. Let F be a family of perfect hash functions from the positions of B to the set C. We consider a function $f \in F$ that associates a distinct color to each of the t positions of B that generates a preserved duo.

We slightly modify the definitions of functions D and P. Define $D[i, C', k']$, for $0 \leq i \leq n$, $C' \subseteq C$ and $k' \leq k$, as a function equal to 1 if there exists a set S_B of $|C'|$ positions of B, each one associated with a distinct color in C', and a set S_A of $|C'|$ positions of $A[1, i]$, such that there exists a mapping from S_A to S_B that preserves k' duos; otherwise, the function is equal to 0. Notice that function D is also defined for the value $i = 0$, although $A[0]$ is not defined, in order to simplify the base case of the dynamic programming recurrence.

$P[h, i, C']$ is a function equal to 1 when there exist positions q and r in B, with $|C'| = i - h$ and $1 \leq q < r \leq |B|$, such that each color in C' is associated with exactly one position between q and r (notice that in [1] each color is associated with exactly one position between q and $r-1$), and substring $B[q, r]$ is identical to $A[h, i]$; otherwise the function is equal to 0.

We can compute $D[i, C', k']$ as follows (exactly as in [1]):

$$D[i, C', k'] = \max \begin{cases} \max_{C'' \subseteq C'} D[h, C'', k''] \times P[h+1, i, C' \setminus C''] \\ \qquad\qquad \text{where } h < i-1,\ i-h-1 = k' - k'' \text{ and } k' - k'' + 1 = |C' \setminus C''| \\ D[i-1, C', k'] \end{cases}$$

In the base case it holds $D[1, C', k'] = 1$ and $D[0, C', k'] = 1$, if $k' = 0$ and $C' = \emptyset$, else $D[1, C', k'] = 0$. There exists a solution of Max-Duo PSM with k preserved duos, if and only if there exists a function $f \in F$ such that $D[n, C, k] = 1$.

Next, we prove the correctness of the recurrence. The proof follow closely that of Lemma 2 in [1], except that here we refer to positions that generate a duo, while in Lemma 2 in [1] we considered a position that induces a duo.

Lemma 1. *Given two related strings A and B, there exists a partial mapping of $A[1,i]$ into B that preserves k' duos generated by positions of B colored by C' if and only if $D[i,C',k'] = 1$.*

Proof. We prove the lemma by induction on i. First consider the base case, that is $i = 0, 1$. Then, $D[i,C',k'] = 1$ if and only if $k' = 0$ and $C' = \emptyset$, since $A[0]$ is a "dummy" position, and $A[1]$ contains no duo.

Now, assume that the lemma holds for $j < i$, we show that it holds for $j = i$.

($\Leftarrow$) First, assume that $D[i,C',k'] = 1$, then we show that there exists a partial mapping of $A[1,i]$ into B that preserves k' duos generated by positions of B colored by C'.

By assuming $D[i,C',k'] = 1$, then, if $D[i-1,C',k'] = 1$, by induction hypothesis there exists a partial mapping of $A[1,i-1]$ into B that preserves k' duos generated by positions of B colored by C'. Assume that $D[i,C',k'] = D[h,C'',k''] \times P[h+1,i,C' \setminus C''] = 1$, for some $C'' \subseteq C'$, with $k' - k'' = i - h - 1$ and $k' - k'' + 1 = |C' \setminus C''|$. Since $D[h,C'',k''] = 1$, by induction hypothesis there exists a partial mapping of $A[1,h]$ into B that preserves k'' duos generated by a set S_B of positions of B colored by C''. Moreover, $P[h+1,i,C' \setminus C''] = 1$, and it follows that there exist positions q and r of B such that each color in $C' \setminus C''$ is associated with a distinct position z of B, with $q \leq z \leq r$, and $B[q,r]$ is identical to $A[h+1,i]$. Hence, it follows that $i - h - 1 = k' - k''$ duos are preserved by mapping $A[h+1,i]$ into $B[q,r]$, and are generated by positions of B colored by $C' \setminus C''$. Notice that, since C'' and $C' \setminus C''$ are disjoint, the sets of positions S_B and $S'_B = \{B[z] : q \leq z \leq r\}$, are disjoint.

As a consequence there exists a partial mapping of $A[1,i]$ into B that preserves k' duos generated by positions of B colored by C'.

($\Rightarrow$) Now, assume that there exists a partial mapping of $A[1,i]$ into B that preserves k' duos generated by positions of B colored by C'. We show that $D[i,C',k'] = 1$. There are two possible cases depending on the fact that positions $i-1$ and i of A generate a preserved duo or not. In the latter case, by induction hypothesis, $D[i-1,C',k'] = 1$ and hence $D[i,C',k'] = 1$.

In the former case, there exists a position h in A, with $1 \leq h \leq i-1$, such that there exists a sequence of preserved consecutive duos $d_{A(h+1,i)}$ mapped into a sequence of preserved consecutive duos $d_{B(z+1,j)}$, with $j - z = i - h$. Since function f assigns a distinct color to each position of B that generates a preserved duo, there exists a set C'' such that each position of $d_{B(z+1,j)}$ that generates a preserved duo with $d_{A(h+1,i)}$ is associated with a distinct color in $C'' \subseteq C'$, and each position of B that generates a preserved duo with a position of $A[1,h]$ is associated with a distinct color of $C' \setminus C''$. Hence, $P[h+1,i,C''] = 1$, for some set $C'' \subseteq C'$. Moreover, by induction hypothesis $D[h,C' \setminus C'',k''] = 1$, with $i - h - 1 = k' - k''$ and $k' - k'' + 1 = |C' \setminus C''|$. By the first case of the recurrence $D[i,C',k'] = 1$. □

From the previous lemma, we can conclude the correctness of the algorithm.

Theorem 1. *Let A and B be two related strings on an alphabet Σ. Then, it is possible to compute if there exists a solution of Max-Duo PSM on instance (A,B,k) in time $(64e^2)^k \text{poly}(n)$.*

Proof. The correctness of the algorithm follows from the correctness of the dynamic programming recurrence (see Lemma 1). Now, we consider the time complexity of the algorithm. The analysis is similar to that of Theorem 1 in [1] by substituting k with t, and observing that, while in Theorem 1 in [1] we considered $D[i,C']$, in this case we consider $D[i,C',k']$. Since $k' \leq k \leq n$, we obtain that the time complexity of the algorithm is $(8e)^t \text{poly}(n)$. Moreover, since $t \leq 2k$, it holds that the overall complexity of the algorithm is indeed $(64e^2)^k \text{poly}(n)$. □

Acknowledgments

We would like to thank Christian Komusiewicz and Mateus de Oliveira Oliveira for pointing out that the FPT algorithm in [1] is incorrect.

References

[1] Beretta, S., Castelli, M., Dondi, R.: Parameterized tractability of the maximum-duo preservation string mapping problem. Theor. Comput. Sci. 646, 16–25 (2016)