Corrigendum to "Bioinformatics Algorithms"

Lemma 4.1.14 states: "After steps 1-3 of phase I, LMS-substrings appear in lexicographic order in A." This is incorrect; the correct formulation of the lemma requires the definition of the so-called LMS-order.

Definition: Let $<$ be a total order on the alphabet Σ. This order induces the LMS-*order* on Σ^* (which we denote by $<_{LMS}$) as follows: For $s, t \in \Sigma^*$, define $s <_{LMS} t$ if and only if either t is a proper prefix of s or $s <_{lex} t$.

It is readily verified that $<_{LMS}$ is a total order. In contrast to the lexicographic order, if an LMS-substring t is a proper prefix of another LMS-substring s, then t is greater then s w.r.t. the LMS-order. For example, the LMS-substrings of $S = acaca\$$ are aca, $aca\$$, and $\$$. We have $aca >_{LMS} aca\$$ but $aca <_{lex} aca\$$.

The corrected version of Lemma 4.1.14 makes use of the total order $\leq_{LMS}$, where $s \leq_{LMS} t$ if either $s <_{LMS} t$ or $s = t$.

Lemma: After steps 1-3 of phase I, LMS-substrings appear in LMS-order in A.[1]

Let $S[i..i+\ell]$ and $S[j..j+k]$ be two LMS-substrings and suppose that position i appears before position j in A (after steps 1-3 of phase I). Then, by the corrected version of Lemma 4.1.14, we have $S[i..i+\ell] \leq_{LMS} S[j..j+k]$. We claim that $S[i..i+\ell] <_{LMS} S[j..j+k]$ implies $S_i <_{lex} S_j$. There are two cases: (a) $S[i..i+\ell] <_{lex} S[j..j+k]$ or (b) $S[j..j+k]$ is a proper prefix of $S[i..i+\ell]$. In case (a), the claim is obviously true. In case (b), the position $j+k$ is S-type (i.e., $T[j+k] = $ S), whereas the position $i+k$ is L-type (i.e., $T[i+k] = $ L). Since the characters $S[i+k]$ and $S[j+k]$ coincide, it follows from Lemma 4.1.8 that S_{i+k} is lexicographically smaller than S_{j+k}. Now $S_i <_{lex} S_j$ is a consequence of $S[i..i+k-1] = S[j..j+k-1]$.

In summary, the lexicographic order of two suffixes that correspond to two different LMS-substrings is clear after steps 1-3 of phase I. The lexicographic order of two suffixes that correspond to two identical LMS-substrings remains to be determined by steps 4-8 of phase I.

---

[1] In fact all LMS-suffixes of length > 1 appear in LMS-order after steps 1-3 of phase I.